

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
Departamento de Química Física I



TESIS DOCTORAL

**Potenciales basados en topología para el estudio de las transiciones de
plegamiento y agregación de proteínas**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Lidia Prieto Frías

Director

Antonio Rey Gayo

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
DEPARTAMENTO DE QUÍMICA FÍSICA I



**POTENCIALES BASADOS EN
TOPOLOGÍA PARA EL ESTUDIO DE LAS
TRANSICIONES DE PLEGAMIENTO Y
AGREGACIÓN DE PROTEÍNAS.**

Memoria que presenta para optar al grado de doctora
en Ciencias Químicas

Lidia Prieto Frías

Director: Antonio Rey Gayo

Madrid, 2008

Agradecimientos.

Esta Tesis doctoral no hubiera sido posible sin la ayuda de Formación de Profesorado Universitario AP2003-0035 del Ministerio de Educación y Ciencia, antes de Educación, Cultura y Deporte, que he disfrutado durante los últimos cuatro años.

Índice general

| | |
|---|-----------|
| 1. Introducción. | 3 |
| 1.1. ¿Qué son las proteínas? | 4 |
| 1.2. Estructura de proteínas. | 8 |
| 1.3. Plegamiento de proteínas. | 9 |
| 1.3.1. Transición termodinámica del plegamiento. | 12 |
| 1.3.2. Topología y embudo de plegamiento. | 15 |
| 1.3.3. Agregación de proteínas. | 18 |
| 1.4. Estudios de plegamiento de proteínas. | 20 |
| 1.5. Potenciales basados en la topología del estado nativo o de tipo $G\bar{o}$ | 22 |
| 1.5.1. Definición de los potenciales de tipo $G\bar{o}$ | 23 |
| 1.5.2. Potenciales basados en topología y termodinámica del plegamiento. . | 27 |
| 1.5.3. Potenciales de tipo $G\bar{o}$ y la agregación de proteínas. | 29 |
| 2. Modelo y método de simulación. | 33 |
| 2.1. Modelo geométrico y de interacción. | 34 |
| 2.2. Método de simulación. | 38 |
| 2.3. Análisis de datos. | 46 |
| 3. Parámetros de un potencial basado en topología. | 57 |
| 3.1. Parámetros del potencial. | 58 |
| 3.2. Mapa de contactos. | 61 |

| | |
|--|------------|
| 3.3. Anchura del potencial. | 66 |
| 3.4. Mapas de contactos con resolución atómica. | 74 |
| 3.5. Estructura de una proteína determinada por diferentes técnicas. | 77 |
| 3.6. Resumen y conclusiones del Capítulo. | 89 |
| 4. Topología y transición de plegamiento de proteínas. | 93 |
| 4.1. Interacciones de largo y corto alcance. | 95 |
| 4.2. Influencia de la rigidez de cadena en la termodinámica del plegamiento. | 97 |
| 4.3. Transición termodinámica de proteínas con distinta topología. | 112 |
| 4.4. Transición termodinámica, topología y rigidez de cadena. | 121 |
| 4.5. Análisis de la topología del estado nativo. | 125 |
| 4.6. Mutaciones del mapa de contactos. | 132 |
| 4.7. Resumen y conclusiones del Capítulo. | 138 |
| 5. Topología y agregación de proteínas. | 145 |
| 5.1. Potencial basado únicamente en la topología del estado nativo. | 147 |
| 5.1.1. Potencial y condiciones de simulación utilizados. | 147 |
| 5.1.2. Sistemas de dos y cuatro cadenas. | 150 |
| 5.2. Modificación de las interacciones locales. | 166 |
| 5.2.1. Potencial y condiciones de simulación utilizados. | 166 |
| 5.2.2. Sistemas de una y dos cadenas. | 171 |
| 5.3. Modificación de las interacciones de largo alcance. | 188 |
| 5.3.1. Potencial y condiciones de simulación utilizados. | 188 |
| 5.3.2. Sistemas de una y dos cadenas. | 189 |
| 5.4. Resumen y conclusiones del Capítulo. | 202 |
| 6. Resumen y conclusiones generales de esta Tesis. | 207 |
| Bibliografía | 213 |

Nadie es una isla

John Donne, "Devotions Upon Emergent Occasions"

Capítulo 1

Introducción.

La vida, tal y como la entendemos, no sería posible sin la existencia de las proteínas. Estas moléculas realizan funciones esenciales en los organismos vivos tales como el transporte del oxígeno en la sangre, el movimiento o el procesado de alimentos. La actividad biológica en sí es un efecto de la actividad proteica. De hecho, organismos unicelulares como las bacterias o las levaduras están compuestas, principalmente, por proteínas y ácidos nucleicos, que son los dos tipos de moléculas que intervienen en los procesos biológicos activos¹. Teniendo en cuenta la importancia de las proteínas, es fácil comprender el gran esfuerzo que desde distintos ámbitos de la comunidad científica se emplean en el estudio de sus propiedades y comportamiento fuera y dentro de los seres vivos.

Una de las características principales de las proteínas es su estructura. La mayoría de estas macromoléculas, en condiciones biológicas, adquieren una estructura tridimensional concreta que determina su actividad biológica. Si no se adquiere esta estructura, la actividad biológica se ve disminuida e, incluso, anulada, pudiendo dar lugar a lo que se conoce como *enfermedades del plegamiento*². La estructura de una proteína en condiciones biológicas viene determinada por la secuencia de aminoácidos, es decir, por los monómeros que la forman y su orden a lo largo de la macromolécula. Conocer cómo se relacionan secuencia y estructura y las propiedades del proceso a través del cual una

determinada secuencia adquiere su estructura funcional es de gran interés en numerosos campos tales como la medicina, la farmacología, la bioquímica y la química-física.

El estudio de este proceso se puede llevar a cabo desde muy distintos puntos de vista, tanto experimentales como teóricos. En este trabajo hemos llevado a cabo un estudio teórico de cómo las características de la estructura funcional de la proteína determinan las propiedades del proceso de adquisición de la misma. Para ello hemos empleado las herramientas de la simulación molecular, utilizando el método de Monte Carlo, combinado con la técnica de *parallel tempering*. Hemos optado por una representación reducida de las proteínas, de manera que el elevado número de grados de libertad del sistema se ve reducido considerablemente. Dado que el objetivo del trabajo es comprobar hasta qué punto la estructura funcional de la proteína determina el proceso de formación de la misma, hemos usado un potencial basado en la topología de esta estructura, también llamado potencial de tipo Gō, cuyas características explicamos más adelante.

1.1. ¿Qué son las proteínas?

Las proteínas son macromoléculas lineales formadas por aminoácidos que se combinan entre sí. Se puede considerar entonces que una proteína es un heteropolímero cuyos monómeros son los aminoácidos. Estos aminoácidos son moléculas orgánicas que contienen un grupo amino ($-NH_2$) y un grupo ácido ($-COOH$). Las proteínas están formadas, casi exclusivamente, por los llamados 20 aminoácidos proteicos. Éstos se caracterizan por tener el grupo amino y el ácido unido al mismo átomo de carbono (llamado carbono α). Este carbono está unido, además, a un átomo de hidrógeno y a un grupo adicional, al cual se denomina cadena lateral y se representa con la letra **R**. Las características de este grupo son las que diferencian unos aminoácidos de otros. En la Figura 1.1 podemos ver la estructura de un aminoácido. Vemos que el átomo de carbono llamado α tiene cuatro sustituyentes diferentes, lo cual le confiere propiedades enantioméricas¹. Las propiedades de los aminoácidos determinan las propiedades de la

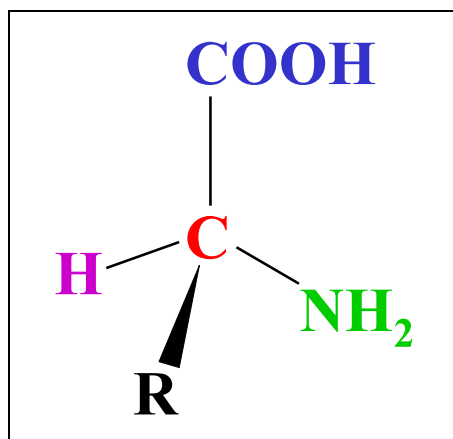
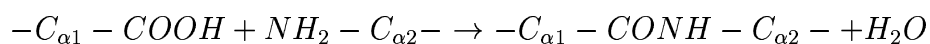


Figura 1.1: Estructura de un aminoácido. **R** corresponde a la cadena lateral y el átomo de carbono de color rojo es el carbono α .

proteína que forman. Dependiendo de la cadena lateral (**R**) de cada aminoácido, éste tendrá unas propiedades diferentes. Una de las clasificaciones posibles en función de las propiedades químicas de los aminoácidos es la que aparece en la Figura 1.2.

Para dar lugar a las cadenas aminoácidas que son las proteínas, estos aminoácidos se unen a través de un enlace llamado peptídico. Este enlace tiene lugar entre los grupos $-COOH$ de un aminoácido y el grupo $-NH_2$ del otro. La reacción que tiene lugar es:



Una característica importante del enlace peptídico es que se puede representar mediante dos formas resonantes, representadas en la Figura 1.3. En una de las formas, el enlace entre el grupo carbonílico y el nitrógeno amínico es sencillo. Sin embargo, en la segunda de las formas, este enlace es doble. Es decir, el enlace peptídico tiene un carácter parcial de doble enlace, lo cual impide la rotación de los grupos unidos a él, que permanecen en un mismo plano (Figura 1.4). Esta característica del enlace peptídico hace que la flexibilidad de la cadena de aminoácidos esté considerablemente reducida y limita su libertad conformacional. Además de las restricciones que impone el enlace peptídico a la libertad conformacional de una proteína, ésta también se ve limitada por el tamaño de las cadenas laterales, que provoca que los valores de los ángulos ϕ y ψ (ver

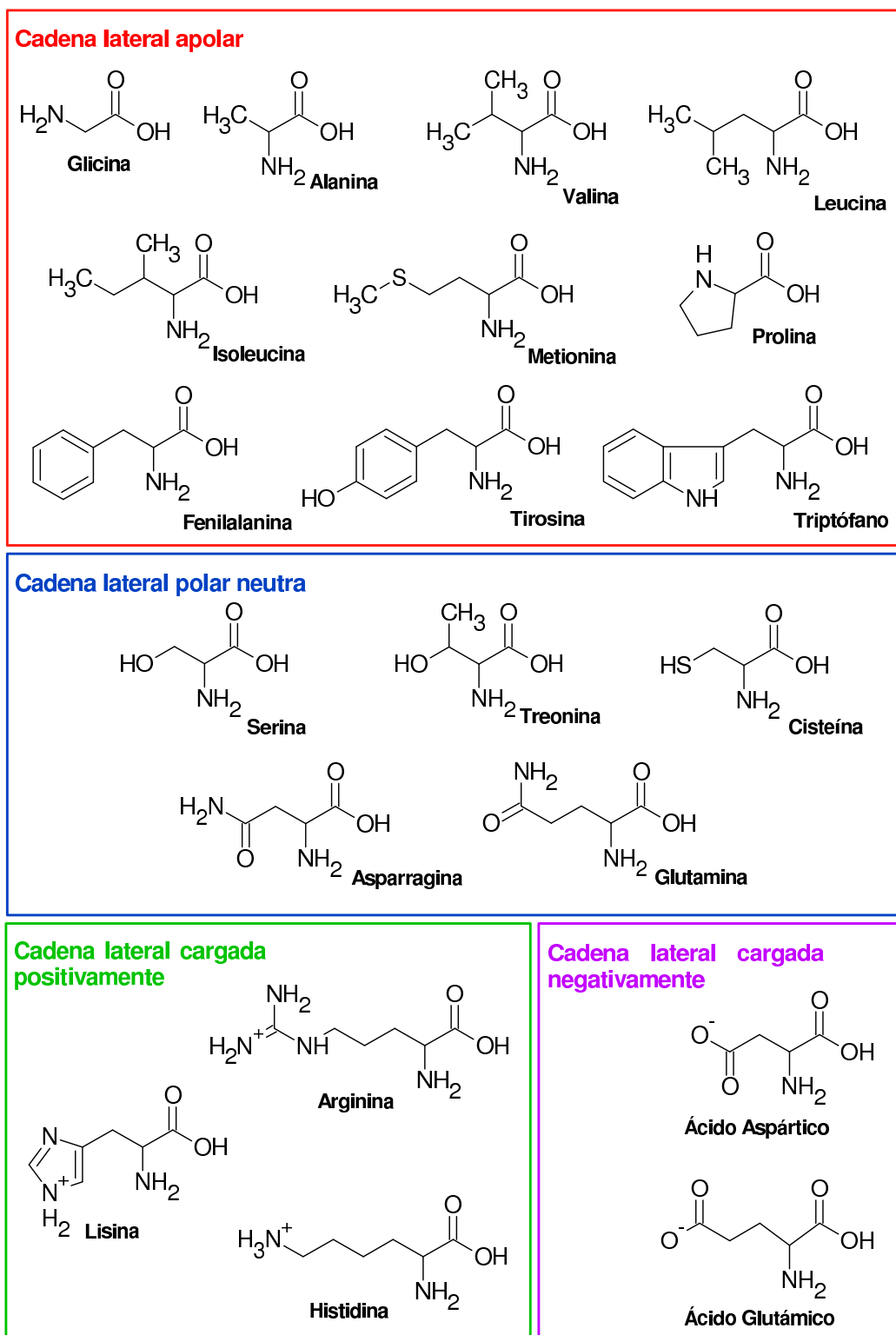


Figura 1.2: Clasificación de los aminoácidos.

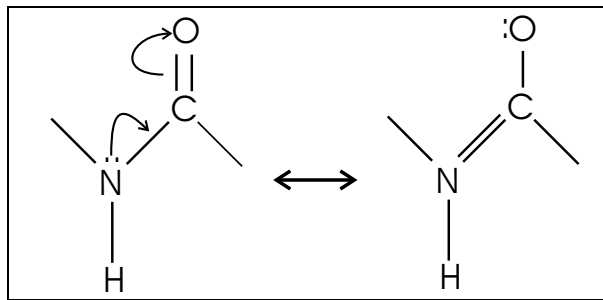


Figura 1.3: Resonancia del enlace peptídico.

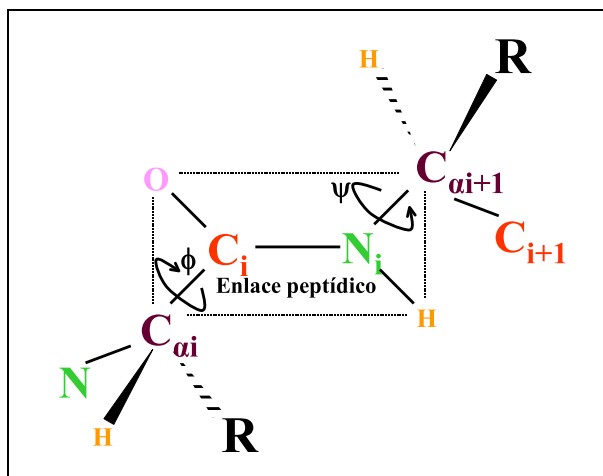


Figura 1.4: Geometría plana del enlace peptídico. Al tener el enlace peptídico carácter parcial de doble enlace, los grupos unidos a él ($C_{\alpha i}$, O, H y $C_{\alpha i+1}$) permanecen en el mismo plano.

Figura 1.4) estén restringidos y adopten únicamente ciertos valores¹.

Por lo tanto, podemos decir que una proteína es una macromolécula compuesta por 20 tipos de aminoácidos distintos. Estos aminoácidos se diferencian unos de otros en su cadena lateral, que determina sus propiedades. Para dar lugar a la cadena de aminoácidos que es una proteína, éstos se unen entre sí a través de un enlace peptídico. Dadas las características de este enlace y de las cadenas laterales de los aminoácidos que componen la proteína, ésta presenta unas propiedades estructurales concretas.

1.2. Estructura de proteínas.

La diferencia entre las proteínas y la mayoría de los heteropolímeros es que, mientras éstos, al polimerizarse, dan lugar normalmente a una distribución de longitudes de cadena y una secuencia de monómeros aleatorias, las proteínas son polímeros lineales no ramificados con una longitud y secuencia de aminoácidos determinadas. Además, son capaces de adquirir una estructura tridimensional concreta. De esta estructura depende la funcionalidad de la proteína. Es decir, estructura y función están íntimamente relacionadas. La secuencia de aminoácidos es la que les confiere a las proteínas la propiedad de asumir su estructura funcional³ o **estado nativo**.

La secuencia de aminoácidos es lo que se conoce como la **estructura primaria de la proteína**. Como ya hemos dicho anteriormente, estos aminoácidos se unen a través de un enlace peptídico cuya geometría impone ciertas restricciones conformacionales a la proteína. Estas restricciones hacen que los ángulos de torsión en torno al átomo de carbono α (Figura 1.4), ϕ y ψ , tomen valores que corresponden a motivos estructurales determinados. De esta forma, con frecuencia los aminoácidos tienden a ordenarse de manera local dando lugar a las conocidas ordenaciones estructurales en forma de hélices α o de cadenas extendidas (láminas β). A estos motivos estructurales de orden local se les conoce como elementos de **estructura secundaria**. Para algunos pares de aminoácidos cuyas cadenas laterales son lo suficientemente pequeñas como para no dar lugar a grandes restricciones conformacionales, los ángulos de torsión pueden tomar valores que no corresponden ni a hélices α ni a láminas β . Esto suele suceder en zonas llamadas lazos o giros, que son más flexibles y permiten la compactación de la proteína haciendo que los elementos de estructura secundaria interaccionen entre sí. A la disposición de estos elementos que da lugar a una estructura tridimensional compacta de la proteína se le llama **estructura terciaria** de la misma¹.

En ocasiones, una proteína está compuesta por varias cadenas de aminoácidos. Éstas tienen una estructura terciaria concreta e interaccionan entre sí dando lugar a la

llamada **estructura cuaternaria de la proteína**¹.

De que la proteína adquiera su conformación nativa depende que pueda realizar su función. Como ya se ha dicho, esta estructura viene determinada por la secuencia de aminoácidos. Al proceso mediante el cual una cadena de aminoácidos adquiere la estructura del estado nativo de la proteína se le conoce como **plegamiento**.

1.3. Plegamiento de proteínas.

La capacidad de las proteínas de adquirir su estructura tridimensional funcional o estado nativo no sólo tiene lugar dentro de los organismos vivos. *In vitro*, estas cadenas polipeptídicas también son capaces de plegarse. De hecho, sucede que si se desnaturaliza una proteína por medio, por ejemplo, de la temperatura, la molécula es capaz de recuperar la estructura del estado nativo cuando se recuperan las condiciones adecuadas. Esto quiere decir que la información respecto a la estructura del estado nativo de una proteína está codificada en la secuencia de aminoácidos⁴. La relación entre estructura y secuencia es directa y está admitida desde hace ya mucho tiempo. Sin embargo, existen diferentes preguntas sobre el plegamiento de proteínas que siguen sin resolver. Un objetivo que se ha perseguido en los últimos años es poder predecir la estructura de una proteína a partir de su secuencia de aminoácidos. Para poder llegar a este objetivo, se han estudiado en profundidad cuáles son las fuerzas que tienen lugar entre los aminoácidos y con el entorno y cómo actúan para llegar a la estructura del estado nativo. Además, está también sin resolver la cuestión de cómo es posible que una proteína pliegue de manera tan rápida dado el elevado número de grados de libertad que posee. Dar respuesta a estas preguntas es lo que se conoce como el *problema del plegamiento*⁵. Solucionar dicho problema supondría la posibilidad no sólo de predecir estructuras a partir de su secuencia, sino de intervenir en el proceso mismo del plegamiento.

En este trabajo nos centramos principalmente en la transición de plegamiento. Probablemente, comprender el mecanismo del plegamiento de proteínas sea la mejor

manera de poder desarrollar un algoritmo computacional que pueda llevar a predecir, mediante principios de tipo físico, la estructura del estado nativo a partir de la secuencia de aminoácidos de una proteína. Por lo tanto, conocer en profundidad la transición de plegamiento es considerado por muchos como el verdadero problema del plegamiento⁵. La primera pregunta a responder es: ¿Se trata de un proceso de búsqueda aleatoria de la conformación de mínima energía o existen fuerzas directrices que determinan caminos de plegamiento? Teniendo en cuenta el elevado número de grados de libertad de las proteínas, el primer caso parece descartado. Si fuera de esta manera, el tiempo de plegamiento sería muy superior al tiempo en el que se produce (del orden de milisegundos). Esto llevó a pensar en la existencia de caminos de plegamiento según los cuales se pasaría de un polímero relativamente flexible (estado desnaturalizado) al estado nativo a través de una secuencia de pasos perfectamente definidos⁶. Por otra parte, el tiempo de plegamiento es normalmente demasiado grande como para pensar en un mecanismo sin barreras para la adquisición de la estructura del estado nativo. Este hecho llevó a una nueva visión del plegamiento de proteínas según la cual, para un mejor entendimiento de este proceso, lo más importante es tener una visión global de su superficie de energía potencial. A partir de estudios cinéticos y teóricos^{7,8} se ha llegado a la conclusión de que la superficie de energía potencial de las proteínas tiene forma de embudo (ver Figura 1.5). En este embudo de plegamiento, la profundidad indica la energía del sistema, de manera que al fondo del embudo, en el mínimo de energía, se encuentra el estado nativo. La anchura del embudo es indicativa de la entropía del sistema. A valores elevados de energía el embudo presenta una mayor anchura que a valores pequeños, lo cual indica que el estado nativo tiene una baja variabilidad conformacional, dado que es un estado único. Sin embargo, el estado desnaturalizado, que es un estado de alta energía, es un estado con una elevada entropía. A pesar de la pendiente negativa que dirige el plegamiento hacia la adquisición del estado nativo, la superficie de energía es rugosa de manera que existen mínimos locales en los que la cadena puede permanecer de manera transitoria⁸.

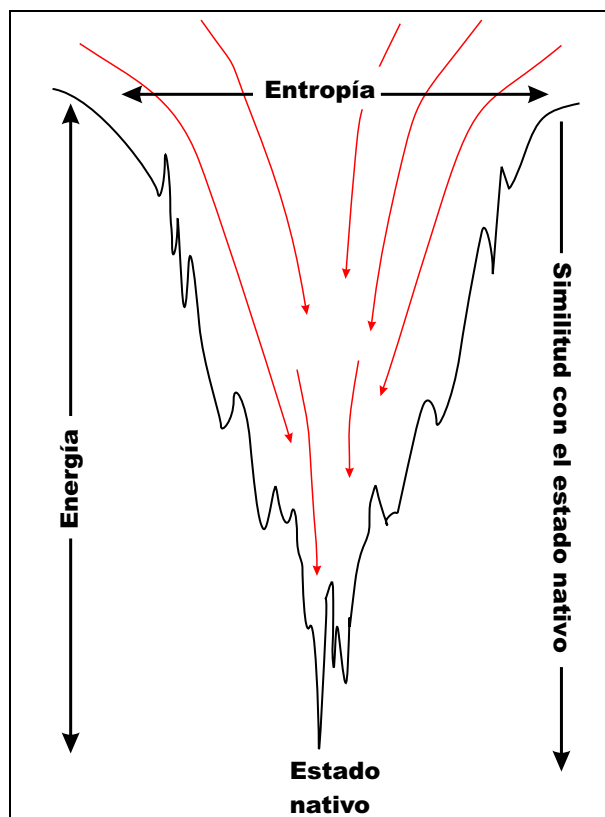


Figura 1.5: Representación esquemática del embudo de plegamiento.

Esta nueva visión del plegamiento de proteínas implica que no existe un único camino de plegamiento, sino varios, a lo largo de la superficie de energía, que dependen de las condiciones del sistema. Esto nos lleva a una visión del plegamiento de proteínas como un proceso de difusión en el que las moléculas individuales disminuyen su energía libre siguiendo diferentes rutas de plegamiento⁹.

La superficie de energía en forma de embudo es lo que distingue a la mayoría de las proteínas del resto de heteropolímeros sintéticos. Esta superficie es el resultado de la evolución, que ha seleccionado secuencias de tal manera que las interacciones presentes en el estado nativo no entren en conflicto unas con otras o con otras interacciones que pudieran existir. Estas interacciones, por el contrario, actúan de manera cooperativa estabilizando la estructura nativa frente a otros estados que corresponden a mínimos locales de energía⁸⁻¹⁰. A esto se le conoce como principio de frustración mínima (*principle*

of *minimal frustration*)⁷. En cualquier caso, si bien es cierto que, como consecuencia de la evolución, las proteínas se consideran sistemas mínimamente frustrados, debido a la presencia de otras interacciones dependientes de la secuencia o de restricciones topológicas, existe siempre cierto grado de frustración. Éste se incrementa con el tamaño de la proteína y su flexibilidad⁸ y proviene, principalmente, de las interacciones no locales.

Podemos decir entonces que, según la visión más actual, la superficie de energía del plegamiento de proteínas tiene forma de embudo mínimamente frustrado. Según esto, el proceso está dirigido a la formación del estado nativo, que corresponde al mínimo de energía del sistema.

Como explicamos a continuación, esta visión tiene unas consecuencias determinadas en el tipo de transición que tiene lugar en el plegamiento y en la relación entre la topología del estado nativo y la forma de la superficie de energía libre del sistema. Veremos, además, las implicaciones que tiene en procesos como la agregación de proteínas.

1.3.1. Transición termodinámica del plegamiento.

Utilizando la teoría de vidrios de espín (*spin glasses*), Bryngelson y colaboradores demostraron que incluir el principio de frustración mínima da lugar a una descripción de la transición de plegamiento como una transición termodinámica de primer orden⁷. Según esto, el paso de un polímero flexible (o estado desnaturalizado) al estado nativo tiene lugar mediante un proceso de dos estados, que resulta en un equilibrio de tipo “todo o nada”. En este tipo de transición, a la temperatura a la que el plegamiento tiene lugar, el perfil de energía libre presenta dos mínimos: uno correspondiente al estado desnaturalizado, favorecido entrópicamente, y otro al estado nativo, que está favorecido energéticamente a través del principio de frustración mínima (ver Figura 1.6). De hecho, está ampliamente aceptado que el plegamiento de las proteínas sencillas de menos de 100 aminoácidos transcurre a través de un proceso de dos estados¹¹. Es decir, se pliegan sin

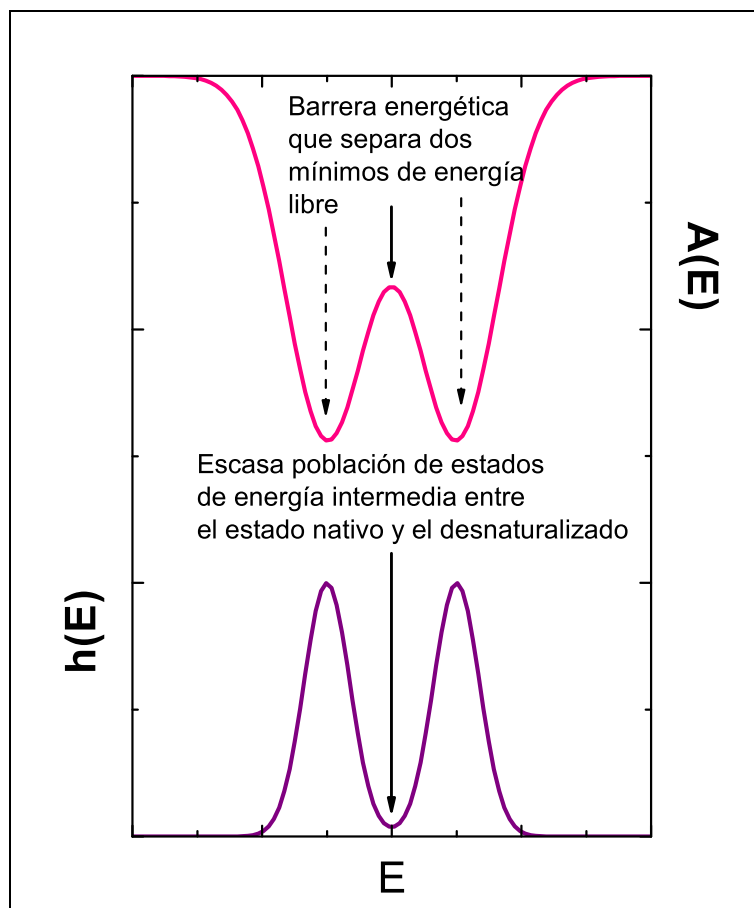


Figura 1.6: Perfil de energía libre ($A(E)$) y población de estados ($h(E)$) para una transición de tipo “todo o nada” a la temperatura de transición. El mínimo de baja energía corresponde al estado nativo y el de alta energía, al estado desnaturalizado.

intermedios en un proceso de tipo “todo o nada”, según el cual la cada cadena de proteína sólo puede encontrarse en uno de los dos estados posibles: nativo o completamente desnaturalizado¹². Proteínas más grandes pueden seguir un proceso de 3 o más estados¹.

Sin embargo, hay proteínas cuyo comportamiento se desvía de este modelo de plegamiento, lo cual puede interpretarse de diferentes formas, dependiendo de las características termodinámicas y cinéticas del sistema dado. En concreto, en el caso de proteínas pequeñas de plegamiento ultra-rápido, en los últimos años se ha postulado la existencia de un plegamiento sin barrera¹³⁻¹⁷. Este tipo de plegamiento daría lugar a un perfil de energía libre con un único mínimo que se desplaza a lo largo de la coordenada

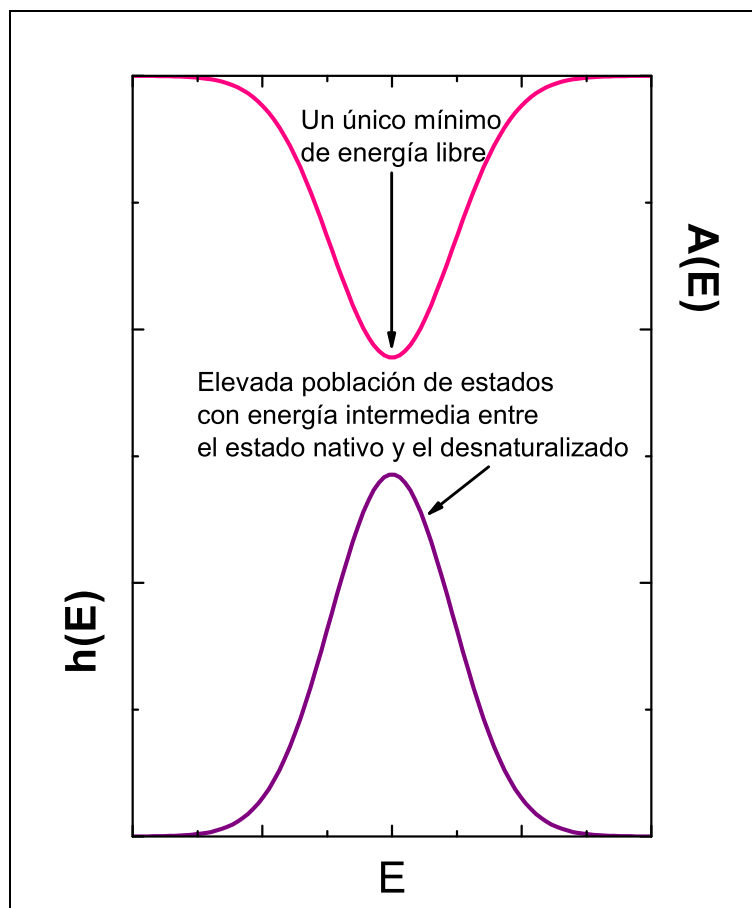


Figura 1.7: Perfil de energía libre y población de estados de una transición sin barrera a la temperatura de transición.

de reacción (plegamiento) con el cambio de temperatura (ver Figura 1.7). En estos casos el proceso tiene lugar de manera continua, implicando un único macroestado a cada temperatura.

Tradicionalmente, el criterio más utilizado y ampliamente aceptado para determinar si una proteína pliega a través de una transición de dos estados es el denominado criterio calorimétrico. A partir de experimentos de calorimetría se puede calcular la entalpía de plegamiento integrando la curva de capacidad calorífica frente a la temperatura. Si esta entalpía es igual o parecida a la llamada entalpía de van't Hoff, que resulta de ajustar la curva de capacidad calorífica a un modelo de dos estados, se considera que la transición es de este tipo¹⁸. En caso contrario se ajusta a un modelo de tres o más

estados. Si bien este criterio se viene aplicando con buenos resultados, parece ser que debe ser aplicado con cautela en el caso de proteínas pequeñas, ya que en estos casos puede ser satisfecho sin necesidad de tratarse de un proceso de dos estados^{13,19}. Además, este criterio parte de la base de que la transición es de dos o más estados, dejando de lado la posibilidad de un plegamiento sin barrera.

Como hemos dicho, la forma de la superficie de energía de las proteínas es consecuencia de la evolución. Por lo tanto, la aparición de una barrera no es un requerimiento químico-físico del sistema, es consecuencia de la selección natural²⁰. Existe la hipótesis de que algunas proteínas podrían haber evolucionado para plegarse de manera ultrarápida y, como hemos dicho, se ha propuesto que algunas de estas proteínas pliegan a través de un proceso sin barreras, lo cual explicaría sus características cinéticas²¹⁻²⁵. Más concretamente, existen estudios que sustentan la hipótesis de que el *complejo multienzimático 2-oxoglutarato deshidrogenasa de Escherichia Coli* [cuyo código del Protein Data Bank (PDB)²⁶ es 1BBL] se pliega mediante un proceso de un único estado^{15-17,22,27}, también conocido como plegamiento tipo *downhill*, como el de la Figura 1.7. Desde luego esta afirmación ha suscitado una gran controversia²⁸⁻³⁴, ya que contradice uno de los paradigmas más aceptados del plegamiento de proteínas.

En resumen, debido a que las proteínas son sistemas con una baja frustración energética, el plegamiento transcurre frecuentemente como un proceso de dos o más estados. Aunque a esta situación se ha llegado a través de la evolución, también existe la posibilidad de encontrar excepciones en las que el plegamiento tiene lugar según una transición de un solo estado.

1.3.2. Topología y embudo de plegamiento.

Si bien es cierto que diferentes proteínas siguen rutas de plegamiento diferentes a lo largo de la superficie de energía³⁵, a partir de mecanismos detallados obtenidos experimentalmente para distintas proteínas se ha observado que muchas de sus carac-

terísticas son similares. Estos patrones comunes se pueden deducir a partir de modelos de embudo perfecto en que la frustración es extremadamente pequeña y la adquisición del estado nativo es inevitable. Un embudo perfecto implica que las únicas interacciones que tienen lugar en el proceso de plegamiento son aquéllas que estabilizan la estructura del estado nativo. Es correcto decir entonces que la topología del estado nativo determina el mecanismo de plegamiento. Dado que la mayoría de las proteínas podrían ser sistemas mínimamente frustrados (aunque no embudos perfectos) se puede considerar la posibilidad de que la topología del estado nativo determine el mecanismo a través del cual se llega a él⁹.

Aunque en muchos casos se conserva cierto grado de frustración^{9,36}, la relación entre mecanismo de plegamiento y topología se ha demostrado experimentalmente. A partir de estudios cinéticos de proteínas pequeñas se ha observado que existe una correlación entre la constante de velocidad de plegamiento (k) y el tamaño de las proteínas²⁴ o su orden de contacto (CO)³⁷ (ver Figura 1.8).

A pesar de que inicialmente se pensó que la velocidad de plegamiento podía estar relacionada con el contenido de hélice α o con el peso molecular, se vio que la correlación con estas variables es pobre. Es la complejidad de la estructura terciaria la que determina la mayor o menor rapidez con que las proteínas adquieren el estado nativo^{38,39}. Esta complejidad se puede medir con el orden de contacto⁴⁰:

$$CO = \frac{1}{NL} \times \sum \Delta Z_{ij} \quad (1.1)$$

donde N es el número de aminoácidos, L es el número de contactos terciarios presentes en el estado nativo y ΔZ_{ij} es la distancia que separa a lo largo de la secuencia a dos aminoácidos i y j que están formando un contacto en el estado nativo. Un valor elevado de CO implica un número de contactos terciarios alto en relación a la longitud de la cadena. Se puede decir que una proteína con mayor CO que otra presenta una mayor complejidad estructural. Como ya se ha dicho y como se ve en la Figura 1.8a, existe una

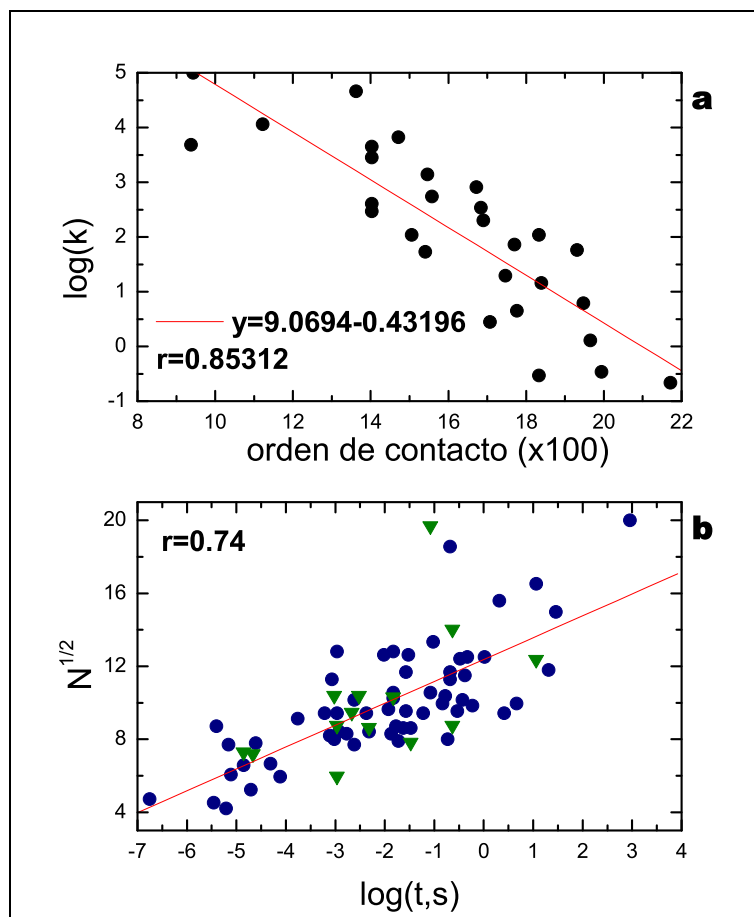


Figura 1.8: Correlación entre estructura y propiedades cinéticas. **a)** Correlación entre la velocidad de plegamiento y el orden de contacto, que cuantifica la complejidad estructural de la proteína (figura adaptada de³⁷). **b)** Relación entre la velocidad de plegamiento y el tamaño de la proteína (figura adaptada de²⁴).

buena correlación entre $\log(k)$ y CO . Aunque el valor de la constante cinética k no se pueda inferir con exactitud a partir del valor de CO , esta relación es importante a la hora de plantear los principios básicos de la teoría del plegamiento de proteínas³⁷.

La superficie de energía está directamente relacionada con la cinética del proceso. Ésta dependerá de la presencia o no de una barrera de energía libre que determine el tipo de transición termodinámica que está teniendo lugar, en función de su altura. Por lo tanto, el perfil de energía libre del plegamiento de una proteína está determinado fundamentalmente por la topología de su estado nativo.

1.3.3. Agregación de proteínas.

En la célula, las condiciones para el plegamiento de proteínas son, hasta cierto punto, poco favorables: la alta concentración de proteína y la presencia de otras macromoléculas que por efecto de volumen excluido aumentan la concentración efectiva de la proteína favorecen su agregación^{1,41}. Por otra parte, cuando el tamaño y la complejidad estructural de la proteína aumentan, el proceso de plegamiento, a su vez, se complica, observándose la aparición de intermedios. En la célula existen mecanismos para evitar los problemas que puedan derivarse de esta mayor complejidad del plegamiento², como por ejemplo la presencia de chaperonas, que son proteínas cuya función es auxiliar el plegamiento de otras proteínas¹. Sin embargo, a pesar de estos mecanismos de protección, en ocasiones se produce un mal plegamiento, dando lugar a estructuras no funcionales o a procesos de agregación². Este mal plegamiento de las proteínas impide que puedan realizar su función, originando las ya mencionadas *enfermedades del plegamiento*. Enfermedades como el Alzheimer, el Parkinson o la enfermedad de Huntington se asocian con la formación de agregados de proteínas. De hecho, los agregados asociados con estas enfermedades tienen la misma estructura² y se les conoce como agregados β -amiloides. Constan fundamentalmente de hebras β que se asocian entre sí de la misma manera que lo hacen en las láminas β características de la estructura de proteínas. Estos agregados pueden crecer bastante, dando lugar a fibras amiloides insolubles. Parece ser que la estructura de las fibras es común para todas las proteínas que dan lugar a este tipo de agregados, independientemente de cuál sea su estructura y secuencia. De hecho, las proteínas que originan agregados β -amiloides observadas *in vivo* presentan una baja homología tanto de secuencia como estructural⁴². Las fibras amiloides que forman diferentes proteínas tienen el mismo difractograma de rayos X, que indica su estructura de tipo β , además de ser largas y no ramificadas, pudiendo alcanzar diámetros de entre 70 y 120 Å⁴³.

Mucho se desconoce todavía sobre la formación de estos agregados y su relación

con las enfermedades a las que se han asociado. Sin embargo, se está haciendo un gran esfuerzo para esclarecer el mecanismo de agregación de proteínas, con el objetivo de intervenir en él para evitarlo y encontrar soluciones a las enfermedades a las que, parece ser, da lugar. Una de las características del proceso de agregación que está más aceptada es la mediación de algún intermedio de plegamiento capaz de asociarse con otra molécula igual⁴³⁻⁴⁶. Bajo condiciones en que la estabilidad del estado nativo está disminuida pero las interacciones no covalentes que tienen lugar no están del todo impedidas, éstas tienen lugar entre moléculas distintas, no dentro de la misma, favoreciéndose así la agregación⁴⁵. Parece probable que las mismas interacciones que tienen lugar dentro de una misma proteína estabilizando el estado nativo sean las que se produzcan entre moléculas diferentes. Siendo esto así, la agregación podría tener lugar a través de un mecanismo de intercambio de dominios (*domain swapping*), siendo también necesaria la presencia de un intermedio de plegamiento que deje expuestos los grupos que interaccionan estabilizando el estado nativo o, en este caso, el agregado⁴⁷⁻⁴⁹. Esta hipótesis es consistente con el principio de frustración mínima y la existencia de un embudo de plegamiento.

Por otra parte, parece ser que la capacidad de formar fibras amiloides es una característica general de las proteínas como heteropolímeros. Es decir, no sólo las proteínas asociadas con enfermedades amiloides son capaces de sufrir este proceso de agregación^{2,45}. Por lo tanto, la formación de estos agregados está producida por interacciones comunes a todas las proteínas. Aunque las superficies de energía de plegamiento de las proteínas son tales que la presencia de intermedios o la posibilidad de agregación y mal plegamiento están minimizadas por acción de la evolución, siempre existe la posibilidad de poblar estados intermedios con población pequeña pero representativa⁵⁰. Una de las formas de evitar la presencia significativa de intermedios de plegamiento que puedan agregar es mediante un mecanismo de plegamiento de dos estados. Es decir, en el momento de producirse el plegamiento, todas las interacciones actúan de manera cooperativa. De esta forma, todas las moléculas de proteína se encuentran en el estado nativo o en el

desnaturalizado, pero no en estados intermedios. Si la estabilidad del estado nativo o la cooperatividad de la transición se ven reducidas por cualquier razón, la agregación está favorecida^{2,51}. Sin embargo, se han observado casos de proteínas que pliegan a través de una transición de dos estados en que el plegamiento ocurre en competición con la agregación. En estos casos, existe un intercambio rápido y reversible entre monómeros desnaturalizados y estados agregados. En otros casos, esta competición no tiene lugar⁴⁸.

En conclusión, puede decirse que, si bien hay propiedades del proceso de agregación que responden al principio de frustración mínima, parece que este proceso está provocado por la competencia entre interacciones no específicas que entran en conflicto unas con otras. A pesar de que las proteínas han sido seleccionadas evolutivamente para impedir su agregación, las condiciones de la célula y la rugosidad de la superficie de energía potencial introducen un grado de frustración en el sistema que favorece este proceso.

1.4. Estudios de plegamiento de proteínas.

Para dar respuestas a las preguntas que plantea el problema del plegamiento, existen dos grandes tipos de tratamientos. Por un lado, el problema del plegamiento se aborda de forma experimental. Dentro de este ámbito, se puede decir a grandes rasgos que hay dos acercamientos posibles: mediante técnicas biofísicas, capaces de monitorizar la variación de una propiedad del sistema a medida que el plegamiento tiene lugar; o a través de la ingeniería de proteínas, analizando las consecuencias cinéticas y termodinámicas de variaciones en la secuencia de aminoácidos². Por otra parte, en los últimos años y gracias al gran avance que han sufrido la ingeniería informática y las técnicas computacionales, se ha desarrollado una vía teórico-computacional para el estudio del plegamiento de proteínas (entre otros problemas)⁵². Dentro de este campo, existen diferentes metodologías y aproximaciones que incluyen desde el uso de algoritmos complejos estudiados con modelos de interacción basados en los principios fundamentales de la química-física, hasta modelos sencillos cuya interacción se define a través de sim-

plificaciones y aproximaciones. En el primero de los casos, se pueden obtener resultados cuantitativos con gran resolución. Sin embargo, este tipo de tratamiento presenta serios problemas relacionados con el tiempo de computación, lo cual hace atractivo el uso de modelos simplificados para el estudio de procesos que ocurren en una escala temporal amplia, como es el caso del proceso de plegamiento. Utilizando este tipo de modelos, los resultados tienen un carácter en ocasiones más cualitativo, pero se pueden obtener conclusiones generales de un problema dado en un tiempo de cálculo más reducido. Cuando se utilizan modelos geométricos y de interacción complejos la relación de los resultados de las simulaciones con los datos experimentales es directa. Sin embargo, si lo que quiere hacerse es un estudio sistemático de las diferentes variables que influyen en las propiedades de un sistema, estos modelos hacen que el estudio sean inabordable, debido al enorme tiempo de computación que requieren. Reduciendo la complejidad del sistema podemos realizar este tipo de estudio en un tiempo razonable sacando conclusiones comparables, al menos cualitativamente, con las observaciones experimentales.

Utilizando un modelo simplificado tanto de la representación geométrica de la proteína como de las interacciones que tienen lugar, en esta Tesis hemos estudiado cuál es la influencia de la topología del estado nativo en las transiciones de plegamiento y agregación de proteínas. De esta forma, se comprobarán los efectos que una superficie de energía potencial en forma de embudo tiene sobre el perfil de energía libre de estos procesos. Se pretende, además, profundizar en la relación entre la estructura de mínima energía y las propiedades del proceso de plegamiento, comprobando hasta qué punto las proteínas son sistemas mínimamente frustrados.

1.5. Potenciales basados en la topología del estado nativo o de tipo $G\bar{o}$.

En los años 70 y 80, N. $G\bar{o}$ y colaboradores desarrollaron el potencial que ahora lleva su nombre⁵³⁻⁵⁵. En estos trabajos originales, el potencial se aplica a modelos en red muy sencillos, trabajando en dos dimensiones. En los modelos en red, cada aminoácido aparece representado por un centro de interacción que ocupa un nudo de una red cuadrada. En esta representación, que no puede originar ninguna estructura realista de una proteína, se define un estado compacto de la “proteína” y se asigna como estado nativo. Dentro de este esquema, se dice que dos aminoácidos están en contacto si son vecinos en la red pero no están enlazados covalentemente. En la Figura 1.9 se puede ver una representación en tres dimensiones de un modelo en red. Pares como el (9-14), marcado en la figura, son los que se considera que están formando un contacto en el estado nativo. Quedan así establecidas cuáles son las interacciones que lo estabilizan y que, según el principio de frustración mínima, son las que dirigen el plegamiento. Las interacciones de tipo $G\bar{o}$, aunque desarrolladas varios años antes que el principio de frustración mínima⁷, definen embudos perfectos de plegamiento⁹. Esto hace posible utilizar estas interacciones para estudiar caminos de plegamiento⁵⁵⁻⁵⁹ y la presencia de intermedios de plegamiento^{60,61}.

Estos potenciales se han utilizado combinados con diferentes técnicas, como pueden ser diferentes variaciones de métodos de Monte Carlo^{53-56,58,62} o de dinámica molecular^{57,59-61}. Aunque en un principio fueron diseñados para ser aplicados a modelos en red, y esto se sigue haciendo^{63,64}, en los últimos años se ha pasado a utilizar modelos fuera de red que permiten una representación más realista de la estructura de una proteína⁶⁵⁻⁶⁷. Incluso se ha aplicado este tipo de potenciales a representaciones atomísticas de las proteínas^{56,57,59,68}. Un modelo fuera de red, en el espacio continuo, complica la definición del potencial de $G\bar{o}$. Así como en un modelo en red se puede determinar claramente

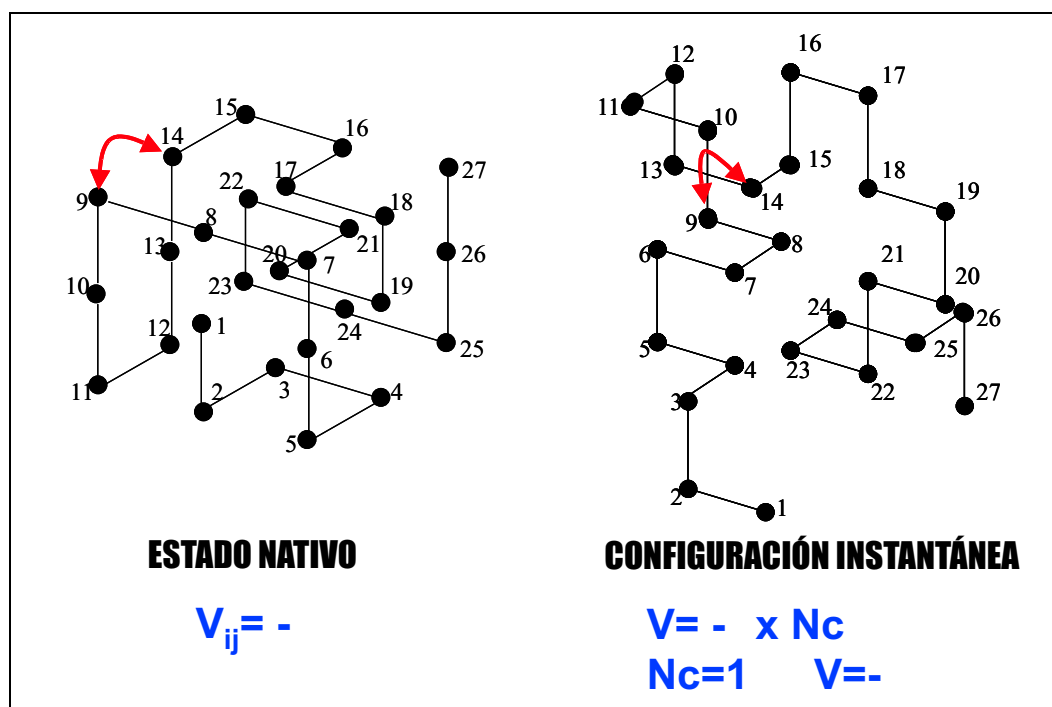


Figura 1.9: Definición del estado nativo con un modelo en red (izquierda) y cálculo de la energía de una configuración instantánea (derecha). N_c es el número de contactos nativos que hay en la configuración instantánea.

qué pares de aminoácidos están interaccionando en el estado nativo, en una descripción continua del espacio esto no se puede hacer de manera igualmente inmediata. Además, el potencial de interacción cambia su forma, ya que debe tener una dependencia espacial continua, al contrario que la descripción original, que es en términos discontinuos. ¿Cuál es entonces la mejor manera de definir el potencial? ¿Cómo afectan los distintos parámetros a la descripción del proceso de plegamiento?

1.5.1. Definición de los potenciales de tipo Gō.

Cuando se aplica el potencial de Gō clásico a un modelo en red, la definición de un contacto es inmediata y el cálculo de energía se hace en términos de la existencia o no de un contacto en una conformación dada. Una vez definido el estado nativo de la proteína se comprueba qué aminoácidos son vecinos en la red (Figura 1.9). Para estos pares de aminoácidos existe una interacción atractiva, mientras que entre los demás no

hay interacción o, dado el caso, se puede considerar que existe una repulsión. La energía de cada contacto en el estado nativo adquiere un valor $-\varepsilon$ y la energía total de este estado es $V_{nat} = -\varepsilon \times Ncon$, siendo $Ncon$ el número de contactos nativos.

A lo largo de una simulación, es necesario calcular la energía de las configuraciones generadas por la misma. Para ello, se cuenta el número de contactos nativos que hay en una configuración instantánea (Nc) y la energía se calcula como: $V = -\varepsilon \times Nc$. De esta forma, el estado nativo es el estado de mínima energía y las únicas interacciones que intervienen en el plegamiento son las que estabilizan el estado nativo.

En el caso de modelos fuera de red el espacio es continuo, lo cual permite describir la estructura de la proteína de manera más realista (Figura 1.10). Pero, como hemos dicho, estos modelos plantean una serie de problemáticas como son la definición de contacto en el estado nativo y la dependencia espacial del potencial de interacción.

Para poder determinar cuáles son las interacciones que estabilizan su estructura en un modelo fuera de red es necesario establecer un criterio para determinar qué pares de aminoácidos están formando un contacto en el estado nativo y cuáles no. Dado que no existen “vecinos” directos, como ocurre con los modelos en red, lo más habitual es establecer una distancia de corte, d_{max} . Aquellos pares que se encuentran a una distancia $d < d_{max}$ en el estado nativo son los que están formando un contacto. Sin embargo, no hay establecido un valor universal apropiado para d_{max} . Definir correctamente su valor es

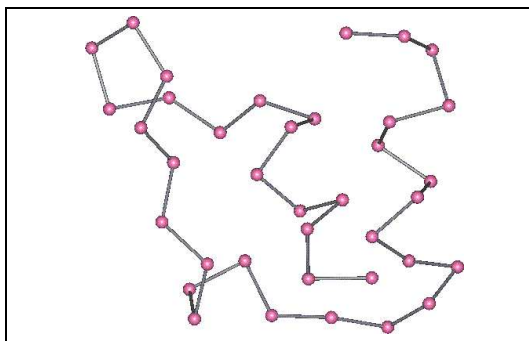


Figura 1.10: Representación geométrica de una cadena polipeptídica con un modelo fuera de red.

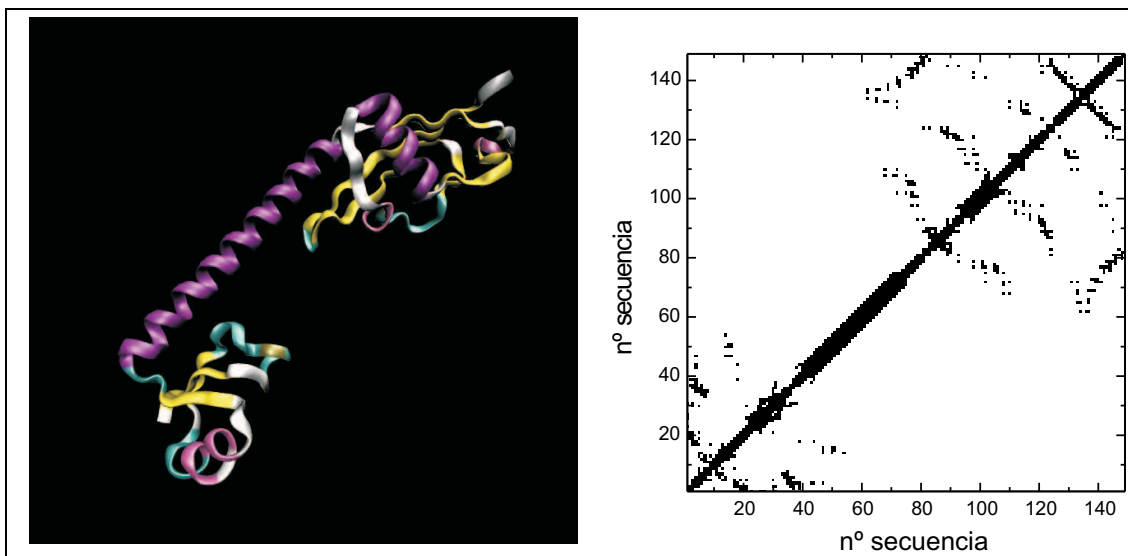


Figura 1.11: Estructura de una proteína (código PDB: 1DVI), realizada con VMD⁶⁹, y su correspondiente mapa de contactos.

de gran importancia para la correcta aplicación de los potenciales basados en topología.

Uno de los modelos fuera de red más utilizados dentro del conjunto de modelos de baja resolución es aquél en que cada aminoácido se representa por un solo centro de interacción que ocupa la posición del carbono α (C_α) de cada aminoácido. Para definir d_{max} , se consideran pues las distancias entre este tipo de átomos. Esto da lugar a una manera alternativa de representar la estructura de una proteína, mediante un mapa de contactos como el de la Figura 1.11. En estos mapas, los ejes corresponden al número del aminoácido a lo largo de la secuencia, y en la coordenada (x, y) se marca un punto en el caso de existir un contacto entre los aminoácidos x e y en la estructura nativa. En los textos clásicos de estructura de proteínas, para hacer este tipo de representaciones se utiliza un valor de distancia de corte entre C_α de $d_{max} = 10$ Å e incluso distancias mayores³. Sin embargo, en los últimos años se han utilizado valores más pequeños. Estos valores varían entre 6.5 y 8.5 Å⁶⁷. En cualquier caso, no existe unanimidad a la hora de elegir un valor adecuado de d_{max} . Probablemente, no haya un valor de d_{max} entre C_α único para todas las proteínas pero, ¿cómo elegir el valor más adecuado?

Por otra parte, una vez definido el valor de d_{max} , ¿cuál es la dependencia espacial

del potencial de interacción? En el caso de los modelos en red se puede decir claramente si un contacto está formado o no y cuál es su energía. En el caso de modelos donde el espacio es continuo, la energía de interacción puede depender de la distancia que separe a un par de aminoácidos. Se hace necesaria, entonces, la definición de una función espacial para el potencial que permita calcular la energía de una configuración determinada. Se han utilizado diferentes funciones analíticas, en muchos casos eligiendo funciones físicas conocidas como pueden ser potenciales de tipo Lennard-Jones^{65,67,70} o de pozo cuadrado^{56,57}. Sin embargo, estas funciones se han tomado como válidas sin presentar un análisis previo de cómo los diferentes parámetros del potencial pueden afectar a la definición de la transición a que dan lugar.

Para dar respuesta a la preguntas planteadas en los párrafos anteriores, en el **Capítulo 3** de esta Tesis hemos realizado un estudio sistemático de cuál es la influencia que tienen los diferentes parámetros de un potencial basado en la topología del estado nativo en las características termodinámicas de la transición de plegamiento. Si bien estas propiedades termodinámicas se han tenido en cuenta en estudios anteriores^{56,71-75}, no se ha llevado a cabo un análisis pormenorizado de los detalles del modelo que, como se verá más adelante, tienen una influencia notable en los resultados de la simulación y en su interpretación. En primer lugar hemos estudiado la importancia del parámetro d_{max} en la definición de un potencial basado en la topología del estado nativo. Aunque su relevancia ya ha sido constatada anteriormente⁶⁷, en este trabajo hemos profundizado en las consecuencias termodinámicas que tiene utilizar un valor u otro. Además, dado que el valor más apropiado de d_{max} entre C_α es dependiente de la proteína, hemos establecido un método para encontrar el valor más adecuado en cada caso. Posteriormente, hemos analizado la influencia de la anchura del potencial. Esta característica determina la correcta descripción del espacio configuracional accesible a un sistema. Para llevar a cabo este análisis, hemos definido el potencial a través de una expresión matemática sencilla que permite controlar de manera fácil y eficaz los parámetros del mismo. Por úl-

timo, determinamos la influencia de las características concretas del mapa de contactos en la definición del proceso de plegamiento.

1.5.2. Potenciales basados en topología y termodinámica del plegamiento.

Según el principio de frustración mínima, la forma de la superficie de energía libre viene determinada por la topología del estado nativo. Esto quiere decir que las características cinéticas y termodinámicas del camino de plegamiento pueden determinarse a partir de las interacciones que estabilizan dicho estado. Como hemos comentado, existe cierta correlación entre la complejidad estructural de una proteína y la cinética de plegamiento y, por ende, también con la termodinámica. Por lo tanto, se pueden utilizar los potenciales de tipo Gō, basados en la topología del estado nativo, para estudiar la transición que tiene lugar.

Una vez aceptado el principio de frustración mínima y definido un potencial de interacción apropiado, la cuestión que se plantea es: ¿qué factores topológicos del estado nativo determinan las características del plegamiento? En general, las respuestas a esta pregunta se han planteado en términos de tipos de interacciones, dividiendo las mismas en interacciones de corto y largo alcance. Mientras que las primeras son aquéllas que tienen lugar entre aminoácidos cercanos en la secuencia, pertenecientes al mismo elemento de estructura secundaria, estabilizando principalmente hélices α , las segundas son las que se producen entre estos elementos, estabilizando la disposición de ellos entre sí y dando lugar a la estructura terciaria de la proteína. Este tema fue tratado en primer lugar por Taketomi y colaboradores en el grupo de N. Gō en el año 1975⁵³. En éste y otros trabajos posteriores, los autores proponen que las interacciones de largo alcance juegan un papel fundamental en el proceso de adquisición de la estructura nativa de la proteína^{53,54}. Por otro lado, existe otra corriente de pensamiento según la cual el estado nativo está determinado principalmente por la acción de las interacciones locales (de

corto alcance)⁴. En un punto intermedio entre estos dos extremos, hay estudios que apuntan a que es necesario cierto equilibrio entre ambos tipos de interacciones para que la transición se produzca de manera cooperativa, tal y como ocurre en la mayor parte de las proteínas^{75,76}.

La contribución relativa de un tipo de interacciones con respecto al otro a la estabilización de la estructura nativa de una proteína está directamente relacionada con el embudo de plegamiento que se define. Aumentar la contribución de las interacciones locales respecto a las terciarias (de largo alcance) tiene diferentes efectos: por un lado disminuye la frustración del sistema, haciendo que el plegamiento esté más dirigido al mínimo de energía libre; por otro lado, el estado desnaturalizado tiene, en este caso, menor energía, desestabilizándose así en términos relativos el estado nativo y, además, habría una mayor diferencia en energía entre los estados no nativos, aumentando así la rugosidad del embudo de plegamiento⁷⁷. El primero de los efectos apunta claramente a que las interacciones locales dirigen el plegamiento^{78,79}. Sin embargo, las otras dos consecuencias de una mayor contribución de las interacciones locales indican lo contrario^{53,54,80-85}.

En relación con la transición del plegamiento, la diferente contribución de unas interacciones frente a otras es un factor determinante de sus características cinéticas y termodinámicas^{22,40,54,75,81,85-94}. Por lo tanto, utilizar potenciales basados en la topología es de gran ayuda ya que, al eliminar la contribución de la secuencia, se puede comprobar cuál es el efecto concreto de los dos tipos de interacciones. De hecho, estos potenciales ya se han utilizado con este objetivo en diferentes trabajos^{54,75,76,78,80,81,85,89,91-93,95}, si bien no se ha profundizado en los motivos termodinámicos concretos por los que tiene lugar un tipo de transición u otro.

Por lo tanto, en el **Capítulo 4** de esta Tesis hemos variado la contribución de los distintos tipos de interacciones de nuestro modelo a la estabilidad del estado nativo. Podemos así estudiar cómo se modifican las propiedades termodinámicas de la transi-

ción dependiendo de la importancia relativa de las interacciones locales respecto a las de largo alcance. Además, hemos seleccionado un conjunto de proteínas que presentan distintas proporciones de estos dos tipos de interacciones en su estructura nativa, lo cual nos ha permitido establecer relaciones entre factores estructurales y el tipo de transición que tiene lugar. Hemos visto que la mayoría de las proteínas pequeñas presentan características estructurales similares que pueden explicar que tengan un comportamiento parecido al adquirir el estado nativo. Por último, hemos profundizado en las características estructurales concretas de determinadas proteínas que pueden explicar su diferente comportamiento, concretamente una transición de plegamiento sin barrera.

1.5.3. Potenciales de tipo $G\bar{o}$ y la agregación de proteínas.

Como ya hemos dicho, parece ser que la agregación se debe a la presencia de cierto grado de frustración del sistema. El hecho de que todas las proteínas puedan formar agregados de estructura similar hace pensar que son las interacciones inespecíficas comunes a todas las proteínas las determinantes del proceso de agregación⁹⁶. Por lo tanto, la agregación es consecuencia de cierto grado de frustración en el sistema. Un potencial basado en la topología del estado nativo es, pues, insuficiente para explicar las propiedades de este proceso. Teniendo esto en cuenta, Ding y colaboradores utilizan variaciones de este potencial añadiendo una perturbación que incluye los puentes de hidrógeno⁹⁶. De hecho, lo más común en estudios teóricos sobre agregación de proteínas en la bibliografía es utilizar potenciales genéricos⁹⁷⁻¹⁰⁶.

Sin embargo, hay casos en que la agregación transcurre a través de un mecanismo de intercambio de dominios. Este mecanismo no viola el principio de frustración mínima y establece que las interacciones que determinan la estabilidad y estructura de los agregados de proteínas son de la misma naturaleza que aquéllas que intervienen en la formación de la estructura nativa de las proteínas¹⁰⁷. De hecho, existe la hipótesis de que el mecanismo de formación de los agregados amiloides es a través de un intercambio de

dominios^{97,108}. En este caso, podemos utilizar un potencial basado en la topología para explicar las propiedades de los agregados de proteínas y su formación. Este potencial se ha utilizado con modelos en red^{109,110} o con sistemas sencillos fuera de red^{111,112} para explicar las características generales de este proceso.

Por otra parte, ya hemos comentado que un mecanismo de plegamiento de dos estados puede ser una manera de evitar la formación de agregados de proteínas. Sin embargo, también existe la hipótesis de que algunas proteínas de plegamiento ultra-rápido podrían haber sido seleccionadas para prevenir su agregación¹¹³. Parece que existe una relación entre el tipo de transición termodinámica y el proceso de agregación de proteínas. Dado que el tipo de transición, como demostramos en esta Tesis, puede estudiarse utilizando un potencial basado en la topología del estado nativo, es posible que sirva también para estudiar el papel de una barrera de energía libre en el proceso de agregación.

Esto nos lleva a estudiar, en el **Capítulo 5**, sistemas multicadena de proteínas utilizando un potencial basado en la topología estableciendo interacciones entre cadenas iguales a las que se producen dentro de cada cadena, como explicamos más adelante. Se pretende comprobar si, así como con este potencial se pueden predecir propiedades termodinámicas del plegamiento de proteínas monoméricas, se puede hacer lo mismo con el proceso de agregación. Dado que ambos están íntimamente relacionados, es de esperar que se puedan explicar, hasta cierto punto, características generales de la competencia entre estos dos procesos. En el caso de no ser así, se plantea la cuestión de qué grado de frustración es necesario para poder explicar dicha competencia. De esta forma queda claro cuál es el límite de aplicabilidad de los potenciales basados en la topología del estado nativo, ya que, como hemos dicho, el proceso de agregación es consecuencia de que siempre existe cierto grado de frustración. Por lo tanto, en este capítulo sólo pretendemos realizar un primer acercamiento al estudio de este proceso, centrándonos en el papel de las propiedades termodinámicas del plegamiento en el proceso de agregación. Además,

queremos destacar cómo las interacciones de tipo general propician en mayor medida la formación de agregados.

No se debe olvidar que, para aplicar los potenciales basados en topología, se debe conocer la estructura del estado nativo. Estos potenciales carecen pues de cualquier carácter predictivo en cuanto a estructura se refiere. La estructura del estado nativo viene determinada por la secuencia, pero el camino a través del cual se adquiere dicha estructura está relacionado con la topología de la misma. Además, es importante recordar que los potenciales de tipo $G\bar{o}$ como el que se utiliza en este trabajo son simplemente una aproximación. Aunque con ellos se pueden explicar propiedades del plegamiento, no sirven para explicar procesos en los que intervienen interacciones no nativas, como puede ser el plegamiento a través de intermedios con estructura muy diferente a la nativa. En este trabajo se pretende estudiar hasta qué punto los potenciales basados en la topología del estado nativo captan los principios básicos del plegamiento, y en ocasiones agregación, de proteínas. Para ello hemos utilizado el método y modelo explicados en el capítulo siguiente.

Capítulo 2

Modelo y método de simulación.

En los últimos años, la simulación molecular se ha convertido en una herramienta científica de gran utilidad. Debido al hecho de que cada vez se diseñan ordenadores y algoritmos más potentes, se pueden simular sistemas que hace no demasiados años eran impensables. Entre estos sistemas se encuentran las proteínas que, debido a su elevado número de grados de libertad, imponen una gran limitación en el tiempo de computación.

La simulación molecular consiste en la reproducción de procesos químico-físicos mediante un programa de ordenador. De esta manera, imponiendo unas reglas de interacción entre partículas, se puede simular el comportamiento del sistema de estudio según esas reglas. Para poder llevar a cabo esta reproducción a nivel microscópico de procesos macroscópicos se necesita una representación geométrica del sistema que se va a simular, un modelo de interacción y un algoritmo de simulación. Dependiendo de las características concretas del sistema de estudio y del objetivo del trabajo que se vaya a realizar, convendrán más unos tipos u otros de modelos y métodos de simulación. Así, si se quieren obtener secuencias temporales de un proceso, como un mecanismo de reacción, es apropiado optar por la Dinámica Molecular. Sin embargo, si se quieren extraer conclusiones de tipo termodinámico como, pongamos por caso, una constante de equilibrio, es más apropiado utilizar el método de Monte Carlo. En cuanto al modelo

geométrico, para realizar un análisis amplio de las diferentes variables de un sistema complejo es más adecuado, como hemos comentado en el Capítulo 1, utilizar una representación simplificada del sistema que permita llevar a cabo el estudio en un tiempo razonable. Un modelo detallado es más apropiado cuando se busca estudiar propiedades concretas de un sistema determinado de manera que se puedan comparar, de forma directa, con resultados experimentales. En el caso del modelo de interacción, si éste parte de los principios básicos de la química-física, no tiene sentido combinarlo con modelos geométricos reducidos que harían que se perdiera toda la exactitud de dicho modelo de interacción. De la misma manera, es, en general, poco útil combinar modelos geométricos muy detallados con modelos de interacción que se basan en aproximaciones o son también una simplificación de las interacciones reales que tienen lugar. Por lo tanto, a la hora de diseñar de forma coherente el modelo y método de simulación, es fundamental tener en cuenta cuál es el objetivo del trabajo que se va a realizar.

2.1. Modelo geométrico y de interacción.

Como ya se ha comentado anteriormente, en este trabajo hemos estudiado las propiedades generales de un potencial basado en la topología del estado nativo y lo hemos utilizado para estudiar cómo ésta determina las propiedades de las transiciones de plegamiento y agregación de proteínas. Para poder llevar a cabo el estudio pormenorizado de las variables del potencial, así como su aplicación a diferentes proteínas y tipos de sistema, todo ello en un tiempo razonable, es necesario optar por una representación simplificada de la proteína. Además, dado que estos potenciales son una aproximación a la realidad y la interacción que definen no responde estrictamente a leyes físicas, la información que se puede obtener con ellos es fundamentalmente de tipo cualitativo. Esto nos lleva a utilizar una función de potencial también sencilla.

Una ventaja fundamental de los modelos reducidos es que requieren un tiempo de computación menor que los modelos más detallados, permitiendo la simulación de

procesos más largos¹¹⁴. En este caso, el modelo geométrico que se ha utilizado consiste en representar cada aminoácido por un único centro de interacción. De esta manera, teniendo en cuenta que los aminoácidos están formados por una media de 20 átomos, se reduce 20 veces el número de grados de libertad. En un modelo con N centros de interacción, el número de interacciones que habría que calcular sería N^N . Si reducimos a $N/20$ el número de centros de interacción, el número de interacciones a calcular se reduce a $(N/20)^{N/20}$. En el caso de una cadena de 5 aminoácidos, con un modelo detallado habría que calcular 100^{100} interacciones, mientras que con el modelo reducido, sólo hay que calcular 5^5 , disminuyendo considerablemente el tiempo de computación. En el modelo que hemos utilizado, el centro de interacción de cada aminoácido se encuentra localizado en la posición del C_α , manteniendo fija la distancia entre centros de interacción consecutivos a lo largo de la secuencia covalente. De esta manera se definen enlaces virtuales de longitud igual a la distancia que separa a dos C_α vecinos en las proteínas reales (3.8 Å). En este modelo las proteínas resultan ser cadenas de cuentas unidas por enlaces virtuales. Cada una de esas cuentas ocupa la posición del C_α del aminoácido. Se ha utilizado, pues, un modelo fuera de red como el que se puede ver en la Figura 1.10.

Entre estos centros de interacción, se ha definido el potencial:

$$u_{ij}(r_{ij}) = \begin{cases} w_{ij} \times [(r_{ij} - d_{ij}^{nat})^2 - a^2] & ; \quad d_{ij}^{nat} - a < r_{ij} < d_{ij}^{nat} + a \\ 0 & ; \quad \text{en el resto de los casos} \end{cases} \quad (2.1)$$

cuya forma se puede ver en la Figura 2.1. En la ecuación (2.1), d_{ij}^{nat} se refiere a la distancia a la que se encuentran los aminoácidos i y j en el estado nativo. Como podemos ver en la Figura 2.1, el potencial tiene la forma de una parábola truncada en $\pm a$. Además, el mínimo de energía corresponde a $r_{ij} = d_{ij}^{nat}$. Cuando todas las distancias entre aminoácidos son iguales a las del estado nativo, la energía es mínima. De esta forma se ha diseñado un potencial sencillo, con el mínimo de energía perfectamente definido en el estado nativo de la proteína: es un potencial basado en la estructura del

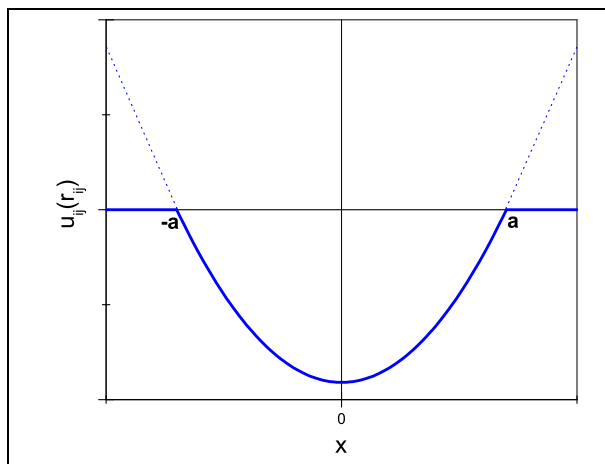


Figura 2.1: Potencial de interacción. $x = r_{ij} - d_{ij}^{nat}$. Las líneas discontinuas indican la forma que tendría el potencial si no estuviera truncado.

estado nativo.

Como ya hemos dicho anteriormente, según el principio de frustración mínima en que se basan estos potenciales, las interacciones que dirigen el plegamiento son aquellas que estabilizan el estado nativo. Por lo tanto, la interacción definida en la ecuación (2.1) tiene lugar únicamente entre aquellos pares de aminoácidos que forman un contacto en el estado nativo. También hemos mencionado antes que la manera de definir qué pares de aminoácidos forman un contacto en el estado nativo viene siendo un tanto arbitraria. En este trabajo hemos clasificado las interacciones que estabilizan el estado nativo en dos grupos: locales y de largo alcance. Las interacciones locales son aquellas que tienen lugar entre aminoácidos cercanos en la secuencia. Dentro de las interacciones locales distinguimos entre *interacciones de ángulo de enlace virtual* e *interacciones de torsión*. Las interacciones de ángulo de enlace virtual son, lógicamente, aquellas que definen el ángulo de enlace virtual, es decir, las que tienen lugar entre el aminoácido i y el $i+2$. Las denominamos también “interacciones 1-3”. Las interacciones de torsión, o “interacciones 1-4”, se refieren a las interacciones entre el aminoácido i y el $i+3$, y definen el ángulo de torsión de tres enlaces virtuales consecutivos. Para incluir la quiralidad de las proteínas en el modelo hemos definido un signo para las distancias $r_{i,i+3}$. Este signo es igual al del

producto triple de los vectores $\mathbf{v}_{i,i+1}$, $\mathbf{v}_{i+1,i+2}$ y $\mathbf{v}_{i+2,i+3}$. En el modelo que se utiliza en este trabajo, se considera que las interacciones locales siempre estabilizan el estado nativo, siendo para ellas $w_{ij} \neq 0$ en todos los casos. El segundo tipo de interacciones que consideramos son las interacciones de largo alcance, también llamadas *terciarias*, ya que son las que estabilizan la estructura terciaria de la proteína. Para determinar cuáles son las interacciones de largo alcance que estabilizan el estado nativo establecemos una distancia de corte, que llamaremos a partir de ahora d_{max} . Si dos aminoácidos i y j ($|i - j| \geq 4$) se encuentran en el estado nativo a una distancia $d < d_{max}$, entonces se considera que i y j están formando un contacto en el estado nativo. Existe, pues, entre ellos una interacción atractiva. Dicho de otra manera, $w_{ij} \neq 0$ sólo en el caso de que los centros de interacción i y j estén a una distancia $d < d_{max}$ en el estado nativo. En caso contrario, $w_{ij} = 0$. Dependiendo de la estructura, una proteína tendrá más o menos contactos distribuidos de una manera u otra. Esto es lo que en este trabajo denominamos *topología del estado nativo*. Una forma de visualizar esta topología es a través de mapas de contactos como el de la Figura 1.11, donde la presencia de un punto indica que existe un contacto en el estado nativo entre el aminoácido i , indicado en el eje de abscisas, y el j , en el eje de ordenadas. El objetivo principal de este trabajo es estudiar cómo esta topología determina las propiedades del plegamiento de la proteína correspondiente.

Aparte del mapa de contactos de la proteína, otras dos variables que intervienen son la profundidad y la anchura a del potencial. Si fijamos $w_{ij} = 1/a^2$ en la definición del modelo de interacciones de la ecuación (2.1) cuando $w_{ij} \neq 0$, el mínimo de energía de cada interacción corresponde a un valor de -1. Por otra parte, el potencial se trunca en un valor a para así poder tener correctamente definido el sistema a elevadas temperaturas. Si el potencial tuviera la forma que define la línea punteada de la Figura 2.1, es decir, si no estuviera truncado, la interacción sería atractiva incluso a distancias muy elevadas. De esta manera, se favorecerían estados colapsados incluso a temperaturas

altas. A estas temperaturas, es de esperar que el conjunto de configuraciones accesibles al sistema sea muy amplio. Si el potencial es atractivo para distancias muy grandes, las configuraciones abiertas se ven desfavorecidas sobre las colapsadas, dando lugar a un estado desnaturalizado de baja entropía compuesto por un número reducido de configuraciones. Sin embargo, truncando el potencial, a partir de cierto valor de distancia la energía se hace cero. De esta manera, el espacio configuracional de elevada energía (con respecto al estado nativo) accesible a temperaturas altas es muy amplio. El valor de la entropía del estado desnaturalizado depende del valor de a . Cuanto menor es la anchura del potencial, mayor es el espacio configuracional accesible al estado desnaturalizado (o estado de alta energía), haciendo que su entropía sea mayor. El valor de a es, por tanto, determinante para una correcta descripción de las propiedades termodinámicas del sistema.

2.2. Método de simulación.

Una vez que se han definido los modelos geométrico y de interacción, es necesario elegir el método de simulación más apropiado. Dado que en este trabajo se estudian propiedades termodinámicas del plegamiento de proteínas, se ha utilizado un método de simulación capaz de reproducir adecuadamente la distribución de estados de equilibrio del sistema a las diferentes temperaturas que se tienen en consideración. Por lo tanto, se ha utilizado el método de Monte Carlo combinado con la técnica de *parallel tempering*^{115,116}.

El método de Monte Carlo consiste en un muestreo aleatorio de la superficie de energía del sistema. Para ello, se lleva a cabo lo que se conoce como un muestreo de importancia de manera que, una vez realizada la simulación, se pueda determinar la probabilidad de encontrar una configuración dada bajo unas condiciones determinadas. A grandes rasgos, una simulación de Monte Carlo consiste en generar configuraciones de manera aleatoria. Cada una de estas configuraciones tiene una probabilidad de aparecer

bajo las condiciones de potencial y temperatura impuestas al sistema. Esta probabilidad viene determinada por el factor de Boltzmann, y para que así aparezca en el muestreo se utiliza el criterio de Metropolis¹¹⁷. Supongamos que partimos de una configuración inicial de energía E_1 . A partir de esta configuración se crea una nueva cuya energía es E_2 .

- Si $E_2 < E_1$ la nueva configuración se acepta y pasa a ser la configuración a partir de la cual se genera una siguiente. La probabilidad estadística de la configuración con energía E_2 es mayor que la de aquélla con energía E_1 .
- Si $E_2 > E_1$, la probabilidad de que se produzca un cambio de un estado de energía E_1 a otro con energía E_2 viene dada por la relación de los factores de Boltzmann de ambos estados:

$$p_{12} = \frac{\exp(-\beta E_2)}{\exp(-\beta E_1)} = \exp[-\beta(E_2 - E_1)] = \exp(-\beta \Delta E) \quad (2.2)$$

donde $\beta = 1/k_b T$, siendo k_b la constante de Boltzmann y T la temperatura. Para aceptar la configuración de energía E_2 con una probabilidad p_{12} se genera un número aleatorio n en el intervalo (0,1). Si $n < \exp[-\beta(E_2 - E_1)]$ el movimiento se acepta y la configuración de energía E_2 pasa a ser aquélla a partir de la cual se genera la próxima. En el caso de que $n > \exp[-\beta(E_2 - E_1)]$, la nueva configuración se rechaza, el sistema permanece en el mismo estado de energía E_1 y éste se considera de nuevo como punto de partida para generar la siguiente configuración.

El método de Monte Carlo se puede esquematizar de la siguiente manera:

- Se genera una configuración inicial y se calcula su energía.
- De acuerdo a una serie de reglas que definiremos más adelante se genera una nueva configuración y se calcula ΔE .

- Se acepta la nueva configuración con una probabilidad igual al mínimo entre $[1, \exp(-\beta\Delta E)]$, según se ha explicado antes.
- Si la nueva configuración se acepta, ésta se contabiliza en la simulación y se repite el proceso a partir de ella, generando una nueva configuración.
- Si la nueva configuración se rechaza, se mantiene la configuración anterior que vuelve a contabilizar en la simulación y se vuelve a tomar como punto de partida para repetir el proceso.

Tradicionalmente, el método de Monte Carlo se ha aplicado a temperatura constante o bajo un esquema de calentamiento o enfriamiento simulado (*simulated annealing*)¹¹⁸. Dicho esquema consiste en variar la temperatura de la simulación gradualmente. El problema de esta metodología es que el sistema tiende a quedarse atrapado en mínimos locales y, por lo tanto, depende fuertemente de las condiciones iniciales y de la velocidad de cambio de temperatura. Para solucionar este problema, en este trabajo hemos implementado la técnica de *parallel tempering* o intercambio de réplicas como alternativa al esquema clásico del calentamiento/enfriamiento simulado.

La técnica de *parallel tempering* consiste en crear N réplicas que evolucionan simultáneamente. Cada réplica o copia se encuentra a una temperatura diferente. Durante un cierto número de ciclos de cálculo, las réplicas evolucionan de manera independiente según el esquema de Monte Carlo anteriormente descrito. En este período, la probabilidad de que se produzca una transición depende únicamente de la temperatura de cada réplica individual. La probabilidad de aceptación y rechazo de las configuraciones que se generan en esta parte de la simulación se calcula según el criterio de Metropolis descrito. Una vez transcurrido un número de ciclos previamente estipulado, se produce un intercambio de configuraciones entre pares de temperaturas consecutivas. Dado que en cada intercambio hay dos copias involucradas, la probabilidad de aceptación de este intercambio depende de la variación en energía y la diferencia de temperatura. Si se

intercambia el estado de energía E_1 que estaba a una temperatura T_1 con el estado E_2 que estaba a una temperatura T_2 :

- A la temperatura T_1 la probabilidad de intercambio $p_{12} = \exp[-\beta_1(E_2 - E_1)]$.
- A la temperatura T_2 la probabilidad de intercambio $p_{21} = \exp[-\beta_2(E_1 - E_2)]$.
- La probabilidad de aceptación de intercambio será por tanto el mínimo entre $(1, p)$, siendo $p = p_{12} \times p_{21} = \exp[-\beta_1(E_2 - E_1)] \exp[-\beta_2(E_1 - E_2)] = \exp[(\beta_2 - \beta_1)(E_2 - E_1)] = \exp(\Delta\beta\Delta E)$

El intercambio de configuraciones es pues una actualización del sistema que disminuye la correlación entre configuraciones sucesivas y mejora así la termalización del sistema a cada temperatura. De esta manera, cada réplica alcanza la distribución de Boltzmann correspondiente a la temperatura a la que se encuentra de manera rápida, evitando en buena medida la posibilidad de quedar atrapado en mínimos locales de energía.

Para garantizar que la probabilidad de intercambio entre dos réplicas sea apreciable, éste debe producirse entre dos temperaturas consecutivas. Los histogramas de energía correspondiente a estas dos temperaturas consecutivas deben tener cierto grado de solapamiento para que el intercambio sea posible (ver Figura 2.2). El conjunto de temperaturas elegido para la simulación debe cumplir que haya un solapamiento significativo entre las distribuciones de energía de temperaturas consecutivas. Además, la mayor temperatura elegida debe ser tal que a esa temperatura se pueda superar cualquier barrera de energía. Esto garantiza que a temperaturas menores se pueda salir de mínimos locales por intercambio con temperaturas mayores. En el caso de la simulación del plegamiento de proteínas llevado a cabo en este trabajo, las características de la transición resultan ser, en muchos casos, de primer orden. Por este motivo, hemos tenido que utilizar un gran número de temperaturas para garantizar un solapamiento adecuado de las distribuciones de energía en la zona de la transición. El solapamiento adecuado de los histogramas de energía se refleja en las fracciones de aceptación de intercambios entre

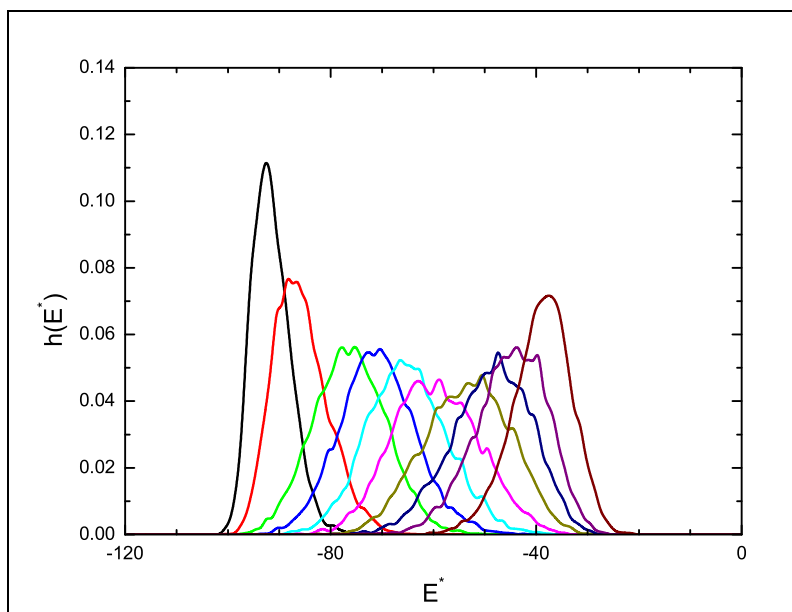


Figura 2.2: Solapamiento de los histogramas de energía. Cada color diferencia la temperatura a la que corresponde el histograma. Datos correspondientes a una de las simulaciones calculadas en este trabajo.

temperaturas. Se considera que un 20 % de aceptación a lo largo de todo el intervalo de temperaturas es suficiente¹¹⁹. Sin embargo, dadas las características de las transiciones simuladas, en este trabajo se ha diseñado el conjunto de temperaturas de tal manera que las fracciones de aceptación se encuentren entre un 40 y un 60 %. Además, para una termalización eficiente es importante que el grado de aceptación se mantenga más o menos constante a lo largo de todo el intervalo de temperaturas¹²⁰.

Se ha demostrado que la técnica del *parallel tempering* permite llegar al mínimo de energía en menos tiempo que el esquema tradicional de calentamiento/enfriamiento simulado. En¹²¹, Gront y colaboradores hacen un estudio comparativo entre los dos métodos viendo que en tiempos de cálculo comparables el método de Monte Carlo combinado con la técnica de *parallel tempering* encuentra el mínimo de energía y, sin embargo, el Monte Carlo clásico no llega a ese mínimo. Además, la variabilidad de los resultados con el Monte Carlo clásico es mayor que con el *parallel tempering*. Esto se debe a la menor capacidad del Monte Carlo clásico de encontrar el mínimo absoluto al tener mayor

tendencia a quedar atrapado en mínimos locales.

El procedimiento que hemos seguido para realizar las simulaciones de este trabajo es el siguiente:

1. En una primera parte se lleva a cabo la generación de la configuración inicial del sistema. En todos los casos partimos de una configuración aleatoria. En el caso de varias cadenas, además, hemos realizado simulaciones previas a altas temperaturas para que la configuración inicial estuviera fuera de cualquier mínimo local.
2. Comienzan los llamados ciclos de Monte Carlo. Cada ciclo consiste en el intento de movimiento de todas las unidades que componen la o las cadenas polipeptídicas. Dentro de un ciclo de Monte Carlo, el movimiento de cada unidad se denomina paso de Monte Carlo. Para generar una nueva configuración en un paso de Monte Carlo llevamos a cabo varios tipos de movimientos:
 - Los extremos de cadena se mueven generando un nuevo vector de enlace virtual de manera aleatoria.
 - Las unidades de interior de cadena pueden moverse según tres tipos de movimientos:
 - Por un lado, generamos una nueva posición del aminoácido i creando un nuevo vector de enlace al azar. La conectividad de la cadena se mantiene desplazando el resto de la cadena una distancia igual al desplazamiento que ha sufrido el aminoácido i (Figura 2.3a).
 - Un segundo tipo de movimiento consiste en la rotación del aminoácido i alrededor del eje que une los aminoácidos $i - 1$ e $i + 1$ como se ve en la Figura 2.3b.
 - También hemos incluido movimientos de pivote. Es decir, un fragmento de la cadena, desde el aminoácido i hasta el final, pivota alrededor del eje x , el y o el z , según se indica en la Figura 2.3c.

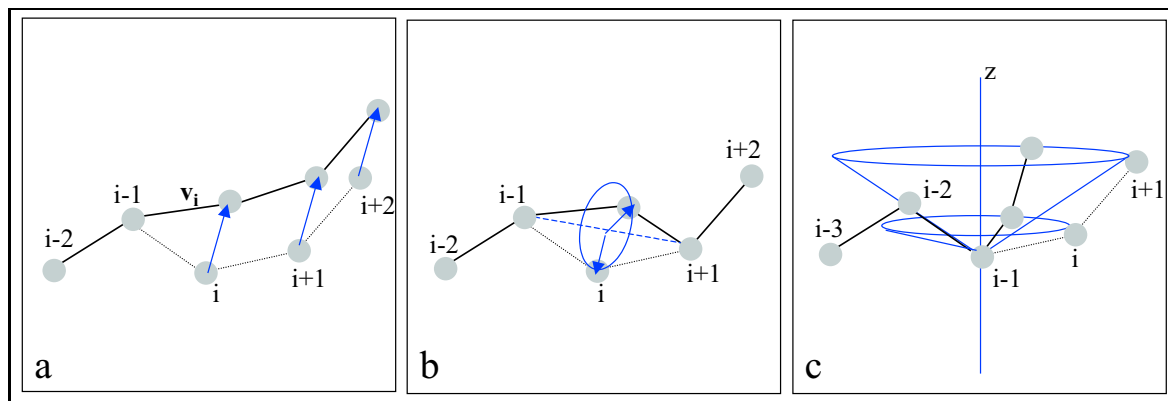


Figura 2.3: Movimientos. En **a)** se genera un nuevo vector v_i para el aminoácido i , el resto de la cadena se desplaza igualmente; en **b)** el aminoácido i toma una nueva posición en el círculo azul y en **c)** se representan los movimientos de pivote alrededor del eje z (azul).

3. Después de cada movimiento lo primero que se hace es comprobar que no existen solapamientos. Consideramos que cada unidad es una esfera rígida de diámetro $1.1 \times bd$ (siendo $bd = 3.8 \text{ \AA}$, la distancia entre C_α). Si hay solapamiento, se vuelve a la configuración anterior y se contabiliza ésta (es decir, se rechaza la nueva configuración). Si no, se pasa al cálculo de energía y se acepta o rechaza el movimiento según el criterio de Metropolis ya explicado.
4. Cuando se ha realizado un número determinado de ciclos de Monte Carlo se lleva a cabo el intento de intercambio de réplicas. El número de ciclos entre un intercambio y el siguiente es una variable de gran importancia para el buen funcionamiento de la técnica de *parallel tempering*. El intercambio de réplicas no debe realizarse de manera demasiado frecuente, ya que el sistema necesita cierto tiempo de relajación para perder la correlación en la nueva temperatura¹²¹. En las simulaciones de este trabajo se han intentado intercambios cada 5000 ciclos de Monte Carlo.

Cada vez que se hace un intercambio, se hace entre todas las temperaturas del conjunto. Es decir, entre todos los pares de temperaturas contiguas. Si se tiene un conjunto de N réplicas a distintas temperaturas: $C = \{T_1, T_2, \dots, T_N\}$, se intenta el intercambio entre las réplicas T_1 y T_2 , T_3 y T_4 y así hasta T_{N-1} y T_N . Cada vez

que se va a hacer el intercambio se elige de manera aleatoria si se empieza en la primera réplica - de manera que el intercambio sea el descrito anteriormente - o si se empieza en la segunda. En este caso, el intercambio se produce entre T_2 y T_3 , T_4 y T_5 , etc... Así, además de evitar que una copia no se intercambie nunca en el caso de que el número de temperaturas sea impar, se favorece la movilidad de las copias a lo largo de todo el intervalo de temperaturas.

El número de réplicas que hemos utilizado depende de las características de la transición. Dado que en muchos casos la transición es de primer orden, para que haya un correcto solapamiento de las distribuciones de energía, el número de temperaturas utilizado es elevado y oscila en nuestras simulaciones entre 15 y 30.

5. Se acepta o rechaza el intercambio de cada par de réplicas según el criterio ya explicado.

Este proceso se lleva a cabo un cierto número de ciclos de Monte Carlo. Primero se lleva a cabo un número de ciclos de termalización. En este período no se registran las propiedades del sistema, dejando que se alcance la configuración de equilibrio a cada temperatura. En este trabajo se han llevado a cabo $2.5 \cdot 10^6$ ciclos de Monte Carlo de termalización. Una vez realizado el equilibrado del sistema, se llevan a cabo $5 \cdot 10^6$ ciclos de Monte Carlo adicionales, a lo largo de los cuales se registran las propiedades del sistema. Para evitar la correlación entre las configuraciones que se registran y obtener un conjunto de configuraciones significativas, el registro de propiedades no se hace después de cada ciclo de Monte Carlo. El intervalo de ciclos de Monte Carlo que hemos usado para la escritura de las propiedades del sistema oscila entre 500 y 1500 ciclos. Para el registro de la estructura se espera un mayor número de ciclos, entre 10000 y 50000. Estos valores dependen de las propiedades del sistema y del análisis que se quiera realizar. Además, para cada conjunto de parámetros estudiados hemos llevado a cabo cinco simulaciones independientes. En cada una de estas simulaciones se parte de una configuración inicial

diferente. De esta manera aumenta la fiabilidad estadística de nuestros resultados, asegurando su reproducibilidad, y nos permite evaluar la desviación típica de las medias de las propiedades estudiadas. Esta desviación aparece representada en forma de barras de error en los resultados que presentamos a lo largo de este trabajo.

Las simulaciones están realizadas a volumen constante, es decir, se está trabajando en el colectivo canónico (NVT)¹¹⁷. En el caso de una única molécula en la que sólo se tienen en cuenta interacciones de la molécula con ella misma, ésta se mueve libremente a lo largo de todo el espacio. Sin embargo, en el caso de sistemas de dos o más moléculas es necesario controlar la concentración. Para ello, se realizan las simulaciones dentro de una caja de simulación con un determinado volumen. En estos casos se pueden tener problemas de superficie, debido a que aquellas moléculas o zonas de una molécula que se encuentran en contacto con las paredes de la caja experimentan interacciones de distinto tipo. Para evitar los problemas que ocasiona la presencia de paredes de la caja, se han aplicado las condiciones periódicas de contorno a través del criterio de mínima imagen¹¹⁷. Esta metodología consiste en dejar evolucionar el sistema de manera que todas las cadenas se muevan libremente. Cuando se quieren calcular las interacciones de la molécula A con la molécula B, se localiza la posición de una molécula B' equivalente a la de B pero dentro de la caja en la que se encuentra A (ver Figura 2.4). Se calculan entonces las interacciones entre A y B'.

Durante la simulación, se registran diferentes propiedades como pueden ser la energía y cada una de sus componentes, el número de contactos nativos de las configuraciones, o valores estructurales como el radio de giro. Estos datos se almacenan en archivos y, para poder estudiar estas propiedades, es necesario realizar el procesado conveniente de los resultados obtenidos.

2.3. Análisis de datos.

A partir de la simulación se obtienen varios ficheros a cada una de las temperaturas:

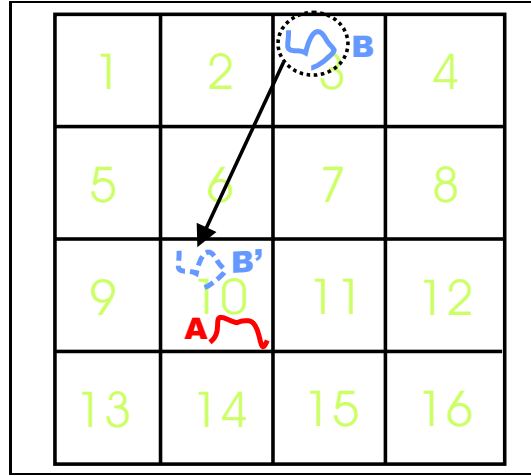


Figura 2.4: Aplicación esquemática del criterio de mínima imagen. Las interacciones entre A y B son equivalentes a las que tienen lugar entre A y B', que se encuentran dentro de la misma caja.

- Por un lado se obtienen datos sobre la energía del sistema. Estos datos aparecen desglosados según las distintas contribuciones a la energía.
- Además, se obtienen datos estructurales como pueden ser el número de contactos nativos que hay presentes en una determinada configuración, el radio de giro o el valor de *rmsd* con respecto al estado nativo.

Rmsd (*root mean square deviation*) es una medida de la diferencia entre dos estructuras. Para calcular el valor de *rmsd* entre dos estructuras se busca la mejor superposición de sus C_{α} ¹²² y después se calcula la desviación cuadrática media entre los átomos equivalentes de las dos estructuras. El valor que resulta es:

$$rmsd = \sqrt{\frac{1}{n} \sum_{i=1}^n (\mathbf{r}_{1,i} - \mathbf{r}_{2,i})^2} \quad (2.3)$$

donde $\mathbf{r}_{k,i}$ es la posición del $C_{\alpha,i}$ en la estructura k y n el número de aminoácidos. Cuanto mayor es esta desviación, más diferentes son las estructuras. A lo largo de nuestras simulaciones hemos registrado el valor de *rmsd* para las distintas configuraciones instantáneas con respecto a la estructura del estado nativo.

Por otra parte, el radio de giro (Rg) se calcula como la raíz cuadrada de la media de la distancia al cuadrado de cada aminoácido al centro de masas:

$$Rg = \sqrt{\frac{1}{n} \sum_{i=1}^n (\mathbf{r}_{\mathbf{CM}} - \mathbf{r}_i)^2} \quad (2.4)$$

siendo $\mathbf{r}_{\mathbf{CM}}$ la posición del centro de masas y \mathbf{r}_i la posición del $C_{\alpha,i}$.

Por último, se calcula el número de contactos nativos en cada estructura a lo largo de la simulación. En el caso de las interacciones de largo alcance, se considera que dos residuos están formando un contacto si se encuentran a una distancia que sólo se diferencie de la nativa en un 10 % o menos. Este cálculo se hace únicamente para aquellos pares que forman un contacto en el estado nativo. En el caso de las interacciones locales (de torsión y ángulo de enlace), la tolerancia es del 5 %.

- Finalmente, se registran también las coordenadas de los átomos para las distintas configuraciones. De esta manera se obtiene una representación de la estructura misma del modelo de proteína.

Además de los valores medios de las diferentes propiedades que se miden a lo largo de la simulación, se pueden también medir otras propiedades que aportan mayor información sobre el sistema. La primera de estas propiedades es la capacidad calorífica a volumen constante (C_v). Si bien lo que se suele medir experimentalmente es la capacidad calorífica a presión constante (C_p), esta diferencia no es significativa, ya que en este trabajo las conclusiones son de tipo cualitativo y, en muchos casos, se trata de un estudio comparativo, por lo que las conclusiones que se sacan a partir de C_v son equivalentes a las que se obtendrían con C_p .

La capacidad calorífica es una de las propiedades más importantes para caracterizar una proteína. Experimentalmente se utiliza para determinar propiedades de la transición así como caracterizar la estabilidad de la proteína (obteniendo datos como la temperatura de transición, T_m , o la variación de entalpía ΔH entre el estado nativo y

el desnaturalizado). La capacidad calorífica tiene muchas definiciones posibles. Se puede interpretar como una medida del aumento de energía del sistema con la temperatura¹²³. Una manera equivalente de definir la capacidad calorífica es como una medida de las fluctuaciones de la energía a una temperatura dada^{13,124}. Cuanto mayores sean esas fluctuaciones, quiere decir que mayor es la variación de energía accesible. A partir de los datos de energía resultantes de las simulación, se puede calcular la capacidad calorífica a volumen constante:

$$C_v^* = \frac{\langle E^{*2} \rangle - \langle E^* \rangle^2}{T^{*2}} \quad (2.5)$$

Cuando tiene lugar una transición de plegamiento se produce un cambio estructural y energético que provoca que las fluctuaciones en la energía accesible a una temperatura dada sean elevadas. Por eso, cuando tiene lugar un tránsito, aparece un máximo en la capacidad calorífica. En la ecuación (2.5), el asterisco indica que las unidades de energía y temperatura en las que trabajamos son unidades reducidas. Los valores de energía que se han utilizado se encuentran en unidades relativas a $k_b T_{ref}$, donde k_b es la constante de Boltzmann y T_{ref} es una temperatura de referencia. De esta forma, la temperatura y la energía que se manejan son magnitudes reducidas (T^* y E^*) y adimensionales, definidas según:

$$T^* = \frac{T}{T_{ref}}$$

$$E^* = \frac{E}{k_b T_{ref}}$$

donde T y E son las magnitudes correspondientes. En la mayoría de los casos, no es posible comparar cuantitativamente los valores absolutos de energía o temperatura de la simulación con resultados experimentales. Sin embargo, sí es posible evaluar cualitativamente la influencia de distintos factores en el proceso de plegamiento. Estableciendo el sistema de referencia, la variación relativa de un parámetro corresponde a una variación equivalente en un sistema real.

Además de la capacidad calorífica, también es importante medir otras propiedades termodinámicas como la energía libre o la entropía del sistema, que caracterizan las propiedades de la transición. Debido a la posible presencia de barreras de energía libre elevadas o de varios mínimos de energía, el cálculo de estas propiedades es bastante complicado bajo un esquema tradicional de simulaciones de Monte Carlo. Sin embargo, el uso de técnicas como *parallel tempering* permite un muestreo mucho más completo de la superficie de energía y, como hemos visto, se evita que el sistema quede atrapado en mínimos locales. Además, el sistema supera las barreras de energía libre con mayor facilidad. Combinando la información obtenida a distintas temperaturas se puede tener una visión global de la distribución de estados energéticos accesibles al sistema. Esto se consigue aplicando el método WHAM (*weighted histogram analysis method*). El método WHAM es una ampliación del método de histograma único¹²⁵ y de histogramas múltiples¹²⁶. En¹²⁵, Ferrenberg y Swendsen propusieron un método de cálculo de propiedades termodinámicas en un amplio intervalo de temperaturas a partir de una simulación a una única temperatura. Haciendo lo mismo pero combinando los resultados de varias temperaturas, desarrollaron el método de histogramas múltiples¹²⁶. A cada temperatura, se exploran regiones distintas del espacio configuracional. Por lo tanto, si se realizan simulaciones en un intervalo amplio de temperaturas, se podrá obtener la distribución de probabilidad de una gran parte del espacio configuracional accesible al sistema y evaluar qué regiones de ese espacio son accesibles a una determinada temperatura. En 1992, Kumar y colaboradores¹²⁷, partiendo de las ecuaciones de Ferrenberg y Swendsen, propusieron el método de histogramas cruzados o WHAM. Con este método, se puede obtener la energía libre a cualquier temperatura y la distribución de energía.

El método WHAM consiste en obtener la mejor estimación de la densidad de estados ($\Omega(E)$)^{127,128}. La densidad de probabilidad de cada estado a cualquier temperatura se puede calcular como:

$$p_m(\beta_k) = Z_k^{-1} \Omega_{mk} e^{-\beta_k E_m} \quad (2.6)$$

donde m indica el estado de energía E_m y Z_k es la función de partición a la temperatura $T_k = 1/k_b\beta_k$. La función de partición se calcula de la manera:

$$Z_k = \int \Omega(E) \exp(-\beta_k E) dE \simeq \sum_{m=1}^M \Omega_m \exp(-\beta_k E_m) \Delta E \quad (2.7)$$

suponiendo que $\Omega_m \cdot \exp(-\beta_k E_m)$ no varía en el intervalo ΔE , utilizado para calcular el histograma de energía.

Por otra parte, la densidad de probabilidad de cada estado puede calcularse a partir del histograma de energía a la temperatura dada como:

$$p_m(\beta_k) = \frac{H_{mk}}{N_k \Delta E} \quad (2.8)$$

donde H_{mk} es el número de configuraciones a la temperatura k que se encuentran en el intervalo de energía ΔE alrededor de E_m , y N_k es el número de configuraciones totales que se recogen de la simulación a esa temperatura k . Igualando las ecuaciones (2.6) y (2.8) se obtiene la expresión para Ω_{mk} :

$$\Omega_{mk} = \frac{H_{mk} Z_k^{-1}}{N_k \Delta E e^{-\beta_k E_m}} \quad (2.9)$$

Dado que la energía libre a la temperatura k se puede expresar como $f_k = \ln Z_k$, la ecuación (2.9) se puede reescribir como:

$$\Omega_{mk} = \frac{H_{mk}}{N_k \Delta E \exp(f_k - \beta_k E_m)}. \quad (2.10)$$

Calculamos así la densidad de estados a una temperatura dada. Sin embargo, lo que interesa conocer es Ω_m , es decir, la densidad de estados total. Como ya se ha dicho, el objetivo es calcular esta función con el menor error posible, es decir, de manera que la incertidumbre estadística sea pequeña, dando más valor a aquellos valores de Ω_{mk} con

menor incertidumbre. Esto lleva a que la densidad de estados es¹²⁸:

$$\Omega_m = \frac{\sum_{k=1}^K g_{mk}^{-1} H_{mk}}{\sum_{k=1}^K g_{mk}^{-1} N_k \Delta E \exp(f_k - \beta_k E_m)} \quad (2.11)$$

siendo g_{mk}^{-1} el valor de la ineficiencia estadística que refleja la correlación entre las configuraciones consideradas. En el caso de simulaciones lo suficientemente largas, cada réplica pasa a través de todas las temperaturas explorando todas las regiones del espacio configuracional. Esto hace que todas las réplicas sean equivalentes y que $g_{m1} = g_{m2} = \dots = g_{mK}$. De esta manera, las ineficiencias estadísticas del numerador y denominador en la ecuación (2.11) se anulan. En la Figura 2.5 se puede ver cómo “viajan” las réplicas por las distintas temperaturas en uno de los sistemas estudiados a lo largo de este trabajo. Tratándose en muchas ocasiones de transiciones de primer orden, el paso de las réplicas de zonas de baja temperatura a las zonas de alta temperatura está dificultado, pero observamos cómo las réplicas visitan todas las temperaturas del intervalo estudiado si hay suficientes temperaturas en la región de la transición. Por lo tanto, en este trabajo se ha considerado que $g_{m1} = g_{m2} = \dots = g_{mK}$. Si bien Chodera y colaboradores consideran que el error en esta aproximación puede ser grande¹²⁸, en el resto de la bibliografía consultada no se hace referencia a la ineficiencia estadística. A la vista de los resultados de este trabajo, se puede decir que el error que se introduce con esta aproximación es pequeño y, en cualquier caso, el mismo para todos los casos, por lo que no afecta a la comparación de los resultados.

En nuestras simulaciones con la técnica *parallel tempering*, el número de configuraciones recogido es igual para todas las temperaturas. Por lo tanto, la ecuación para

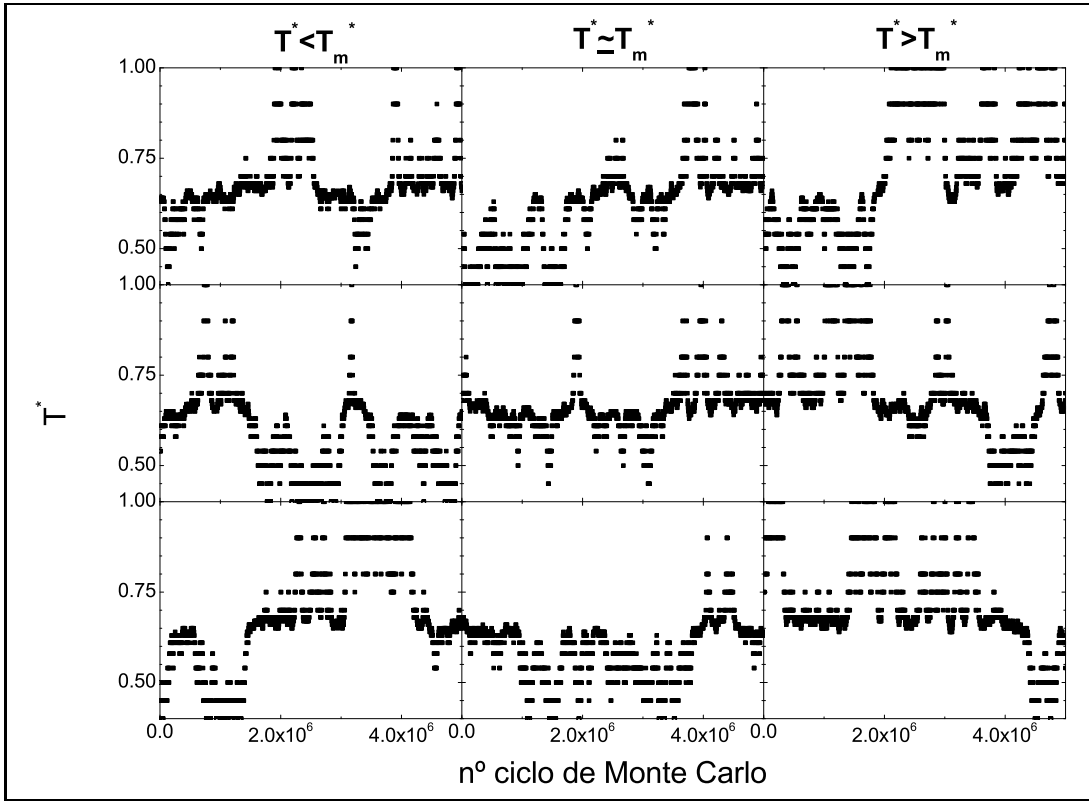


Figura 2.5: “Viaje” de las distintas réplicas por las distintas temperaturas. Cada cuadrado corresponde a una réplica. En el eje de ordenadas se representan las temperaturas por las que pasan las distintas réplicas. Las tres réplicas de la columna de la izquierda son aquéllas que al principio de la simulación estaban a $T^* < T_m^*$, las tres de la columna del centro parten de una temperatura inicial cercana a T_m^* y las réplicas de la columna de la derecha, están, al principio de la simulación, a una temperatura $T^* > T_m^*$.

obtener la densidad de estados queda como:

$$\Omega_m = \frac{H_m}{N \Delta E \sum_{k=1}^K \exp(f_k - \beta_k E_m)} \quad (2.12)$$

siendo H_m el número total de configuraciones en el intervalo de energía ΔE alrededor de E_m , y N el número de configuraciones recogido a cada temperatura. Sabiendo que:

$$f_k = -\ln \sum_{m=1}^M \Delta E \Omega_m e^{-\beta_k E_m} \quad (2.13)$$

los valores de Ω_m y f_k se pueden calcular a partir de los datos de la simulación por resolución iterativa de las ecuaciones (2.12) y (2.13), partiendo de un valor inicial de $f_k = 0$ para todas las temperaturas¹²⁷, hasta alcanzar la autoconsistencia.

El valor de ΔE es importante para el error en el cálculo de Ω_m . Cuanto mayor es ΔE , mayor es este error. Por otra parte, un valor demasiado pequeño también hace que el error crezca. En este trabajo, $\Delta E = 1$.

Una vez estimada de la mejor manera posible la densidad de estados del sistema, se puede calcular cualquier propiedad termodinámica en el intervalo de temperaturas simulado. Se pueden calcular valores medios de estas propiedades:

$$\langle O \rangle(T) = \frac{\sum_{m=1}^M O_m \Omega_m e^{-\beta E_m}}{\sum_{m=1}^M \Omega_m e^{-\beta E_m}} \quad (2.14)$$

así como diferencias entre dos estados. Poniendo por caso que el estado nativo corresponde a un intervalo de energía $E_m = (E_1, E_l)$ y el estado desnaturalizado al intervalo $E_m = (E_{l+1}, E_M)$, en un equilibrio de dos estados, la diferencia de la propiedad O entre esos dos estados se puede calcular como:

$$\Delta O(T) = \frac{\sum_{m=l+1}^M O_m \Omega_m e^{-\beta E_m}}{\sum_{m=l+1}^M \Omega_m e^{-\beta E_m}} - \frac{\sum_{m=1}^l O_m \Omega_m e^{-\beta E_m}}{\sum_{m=1}^l \Omega_m e^{-\beta E_m}} \quad (2.15)$$

Con la función de densidad de estados también se puede calcular la entropía de cada estado como:

$$S_m = \ln \Omega_m \quad (2.16)$$

y dado que la energía del estado m es E_m , se puede calcular la energía libre de Helmholtz del estado m según la ecuación:

$$A_m(T) = E_m - T \times S_m. \quad (2.17)$$

Representado $A_m(T)$ frente a E_m se obtiene el perfil de energía libre a la temperatura T . En este caso, se representa la energía libre de un estado frente a la energía interna de dicho estado, pero puede interesar representar la energía libre de un estado frente a otra propiedad del mismo. Para ello, debe hacerse el cálculo del método WHAM utilizando histogramas bidimensionales o tridimensionales¹⁰⁵. De esta forma se obtiene la densidad de estados y la energía libre como funciones de parámetros distintos a la energía. Siendo ξ_1 y ξ_2 dos propiedades cualesquiera del sistema, se obtiene:

$$\Omega_m(\xi_1, \xi_2) = \frac{H_m(\xi_1, \xi_2)}{N \Delta E \sum_{k=1}^K \exp(f_k - \beta_k E_m)} \quad (2.18)$$

$$f_k(\xi_1, \xi_2) = -\ln \sum_{m=1}^M \Delta E \Omega_m(\xi_1, \xi_2) e^{-\beta_k E_m} \quad (2.19)$$

Para calcular el perfil de energía libre frente a los parámetros ξ_1 y ξ_2 a una temperatura dada, se calcula:

$$A(\xi_1, \xi_2) = -k_b T \ln \sum_{m=1}^M \Omega_m(\xi_1, \xi_2) e^{-\beta E_m} \quad (2.20)$$

De esta manera se pueden obtener perfiles de energía libre mono y bidimensionales frente a propiedades estructurales del sistema.

Con la descripción del cálculo de todas estas propiedades del sistema, se puede pasar a estudiar la termodinámica del proceso de plegamiento utilizando los potenciales de tipo Gō.

Capítulo 3

Parámetros de un potencial basado en topología.

Como hemos comentado en el Capítulo 1, los potenciales basados en los contactos presentes en el estado nativo se utilizan con frecuencia en simulación molecular para predecir características dinámicas, termodinámicas y estructurales del proceso de plegamiento. Sin embargo, la forma matemática que se da al potencial se elige, en general, de forma arbitraria y no se habían estudiado de manera sistemática, hasta ahora, las implicaciones que pueda tener en la descripción de la transición de plegamiento. En este capítulo, utilizando el modelo y método descritos en el apartado anterior, analizamos las consecuencias que los distintos parámetros del potencial tienen en las propiedades termodinámicas del plegamiento de proteínas. En una primera parte observamos que la definición del valor de d_{max} determina la presencia de una transición más o menos compleja, con o sin intermedios termodinámicos en el plegamiento del modelo. Además, comprobamos cómo la anchura del potencial es un parámetro crucial para la presencia o no de barrera de energía libre entre el estado nativo y el desnaturalizado. Esto se debe, fundamentalmente, a la influencia que este parámetro ejerce sobre la entropía del estado desnaturalizado¹²⁹. Finalmente, estudiamos cómo la topología general del estado nativo

determina las propiedades de la transición en este tipo de modelos.

3.1. Parámetros del potencial.

Como se ha visto en el capítulo anterior (ecuación (2.1)), el potencial de atracción entre dos residuos que forman un contacto en el estado nativo tiene la forma:

$$u_{ij}(r_{ij}) = \frac{1}{a^2} \times (x^2 - a^2) \quad (3.1)$$

para $x^2 < a^2$, donde $x = r_{ij} - d_{ij}^{nat}$, en caso contrario $u_{ij}(r_{ij}) = 0$. En este trabajo hemos dado distintos valores al parámetro a para estudiar cómo la anchura del potencial de atracción influye en la termodinámica del plegamiento de este modelo de proteínas. Un valor de a pequeño implica un potencial de atracción estrecho (ver Figura 3.1). Sólo aquellos pares de residuos que en una conformación cualquiera se encuentren a distancias muy próximas a la nativa contribuyen significativamente a la energía. Por el contrario, valores altos de a implican potenciales anchos. Con este potencial, una determinada configuración tiene normalmente menor energía (más negativa) que con el potencial más estrecho, ya que hay un mayor número de pares que contribuyen a la misma (ver Figura 3.1). Los valores que se han dado a este parámetro varían entre 0.3 y 1.0 Å, como se recoge en la Figura 3.1.

Por otra parte, se ha estudiado la influencia de d_{max} en la termodinámica de la transición de plegamiento. Este parámetro define qué contactos son los que estabilizan el estado nativo de la proteína. Cuando dos residuos se encuentran a una distancia $d < d_{max}$ en el estado nativo, se considera que están formando un contacto. Dado que el modelo que se está utilizando es un modelo reducido a los C_α de la proteína, consideramos, inicialmente, un valor de d_{max} entre este tipo de átomos. Como ya hemos explicado en el Capítulo 1, en este trabajo hemos estudiado dos valores de este parámetro: el valor clásico de 10 Å? una distancia más corta de 7.5 Å, que se encuentra dentro del intervalo

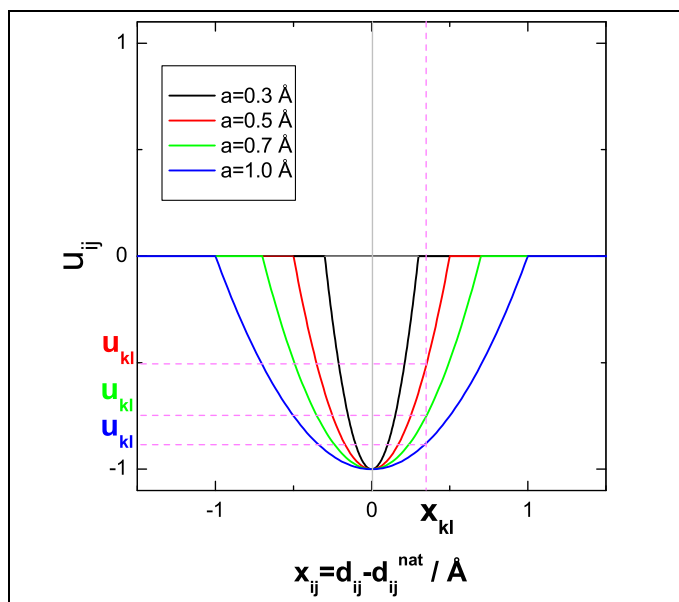


Figura 3.1: Forma del potencial para los diferentes valores de a .

de distancias más utilizadas en trabajos recientes^{3,67}.

Para realizar este estudio, hemos elegido una proteína cuya transición ha sido ampliamente caracterizada desde un punto de vista termodinámico. Esta proteína es el dominio B1 de la proteína G de *Streptococcus* [cuyo código del Protein Data Bank²⁶ (PDB) es 2GB1¹³⁰, con el cual se denominará a la proteína en adelante]. La estructura, que se puede ver en la Figura 3.2, consta de dos horquillas β que forman una lámina β sobre la que se dispone una hélice α . En la misma figura aparecen los mapas de contactos que caracterizan la estructura, definidos con una distancia $d_{max} = 10$ Å (a la izquierda en la figura) y con $d_{max} = 7.5$ Å (a la derecha en la figura). Ambos mapas recogen las características generales de la estructura: los contactos entre la hebra 1 (aminoácidos del 1 a 8) y la hebra 2 (13-20) que estabilizan la horquilla 1; los contactos entre la hebra 3 (42-46) y la hebra 4 (51-56) que aparecen en la horquilla 2; los contactos locales que dan lugar a la hélice α (22-37) y los contactos que estabilizan la lámina (entre la hebra 1 y la hebra 4). Esta proteína tiene 56 aminoácidos, es decir, es una proteína pequeña, lo que la hace especialmente apropiada para estudios computacionales. Además, ha sido objeto de numerosos estudios tanto experimentales^{131–135} como teóricos^{68,136–138}.

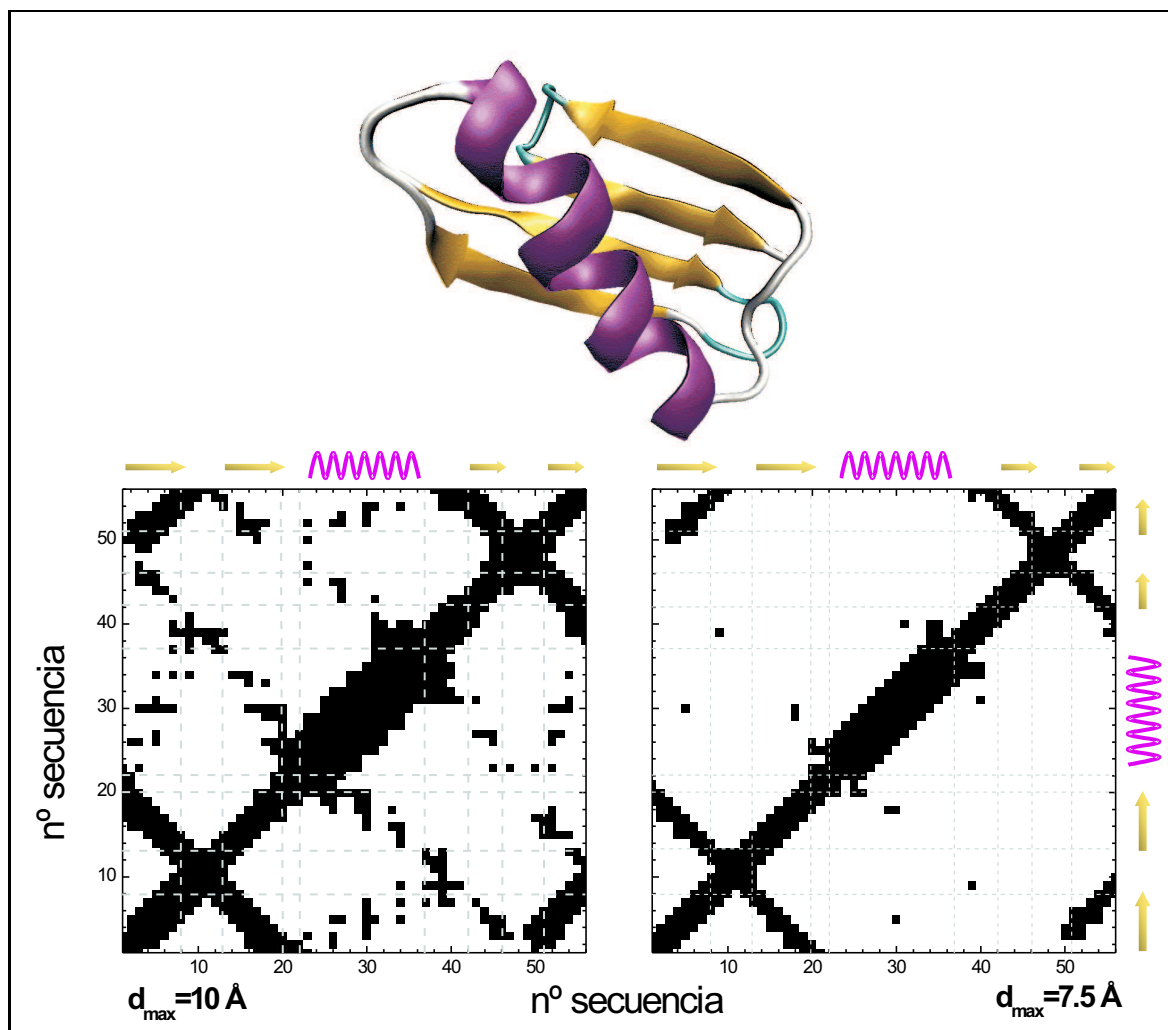


Figura 3.2: Estructura de 2GB1 (realizada con VMD⁶⁹) y el mapa de contactos que se obtiene con $d_{max} = 10 \text{ \AA}$ (izquierda) y con $d_{max} = 7.5 \text{ \AA}$ (derecha). Las líneas discontinuas separan los distintos elementos de estructura secundaria.

Las principales conclusiones de estos estudios que interesan aquí son que esta proteína pliega de manera cooperativa, a través de un proceso de dos estados, sin intermedios termodinámicos detectados, si bien se han caracterizado algunos intermedios cinéticos. Desde un punto de vista termodinámico, la transición que tiene lugar es de primer orden. Dada la estabilidad de esta proteína, se puede decir que la barrera de energía libre presente entre el estado nativo y el desnaturalizado es bastante grande.

3.2. Mapa de contactos.

Como se ha explicado en el capítulo anterior, el mapa de contactos es una representación gráfica de la estructura de una proteína (ver Figuras 1.11 y 3.2). En estos mapas aparecen representados aquellos contactos que estabilizan el estado nativo. Es decir, éstos son los pares de aminoácidos a los cuales se les asigna un valor $u_{ij} \neq 0$. Dependiendo del valor de la distancia de corte (d_{max}), el mapa es diferente. El mapa de contactos que se obtiene utilizando un valor $d_{max} = 10$ Å considera todos los contactos que aparecen en el mapa definido por $d_{max} = 7.5$ Å, más todos aquellos pares de aminoácidos cuyos C_α se encuentren a una distancia entre $(7.5, 10]$ Å.

Se han llevado a cabo varias simulaciones independientes con el potencial definido de acuerdo a cada uno de los mapas de la Figura 3.2 para los cuatro valores de a de la Figura 3.1. Los resultados de estas simulaciones nos permiten estudiar el tipo de transición que se obtiene con los distintos parámetros del potencial, calculando la capacidad calorífica del sistema a distintas temperaturas. En la Figura 3.3 se pueden ver los resultados de este cálculo. En todos los casos, la presencia de un pico en la capacidad calorífica es indicativa de la transición entre el estado nativo y el desnaturalizado. Comparando los resultados obtenidos en el caso de $d_{max} = 7.5$ Å y $d_{max} = 10$ Å, lo primero que se observa es que estos valores tienen un efecto drástico sobre el pico de capacidad calorífica. Mientras que para un valor de 10 Å la capacidad calorífica presenta un único pico, es decir, una única transición, cuando se utiliza un valor de 7.5 Å aparecen dos picos. Estos dos picos, que a valores elevados de a se distinguen como un pico y un hombro en la capacidad calorífica, se diferencian perfectamente para valores pequeños de dicho parámetro. Cada uno de estos picos corresponde a una transición diferente, lo cual señala la presencia de un intermedio termodinámico de plegamiento cuando se utiliza el mapa que se obtiene con esta distancia de corte.

A partir de las configuraciones recogidas a temperaturas intermedias entre los dos picos, hemos comprobado que los resultados de la simulación muestran estructuras en

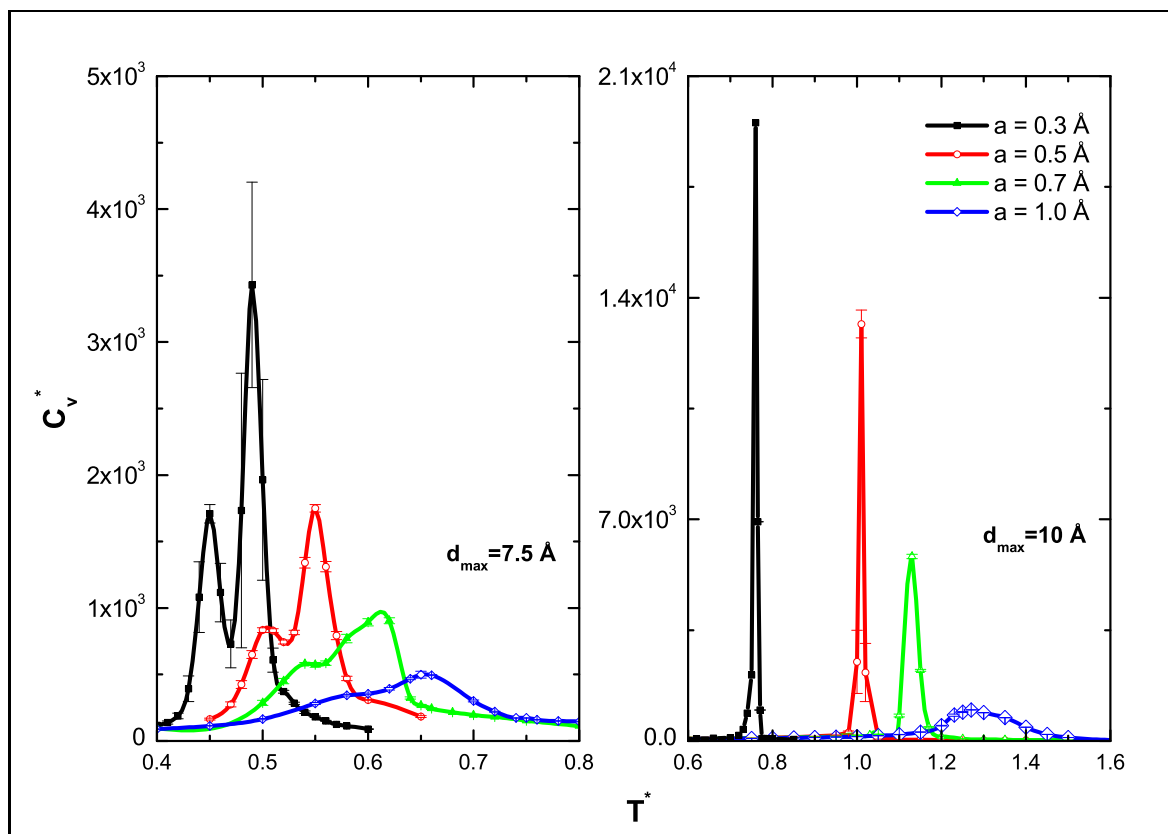


Figura 3.3: Variación de la capacidad calorífica a volumen constante frente a la temperatura. A la izquierda, los resultados obtenidos con $d_{\max} = 7.5 \text{ \AA}$ y a la derecha con $d_{\max} = 10 \text{ \AA}$. Cada color, indicado en la leyenda, corresponde a los resultados obtenidos para un valor de a diferente.

las que la lámina β se ha formado mientras que la hélice no presenta una estructura definida. Para poder representar esta observación hemos calculado la frecuencia con que un contacto nativo está formado en el conjunto de configuraciones que aparecen a la temperatura del mínimo de capacidad calorífica (ver Figura 3.4). Para calcular dicha frecuencia hemos contabilizado en cuántas de todas las configuraciones que se registran a esa temperatura aparece formado cada uno de los contactos nativos. Vemos que los contactos que estabilizan la hélice α aparecen en color más claro. Esto quiere decir que, en las estructuras recogidas a la temperatura estudiada, estos contactos aparecen con menos frecuencia que el resto. Sin embargo, los contactos que estabilizan la lámina β aparecen con una frecuencia mucho mayor, indicando que, a esta temperatura, la lá-

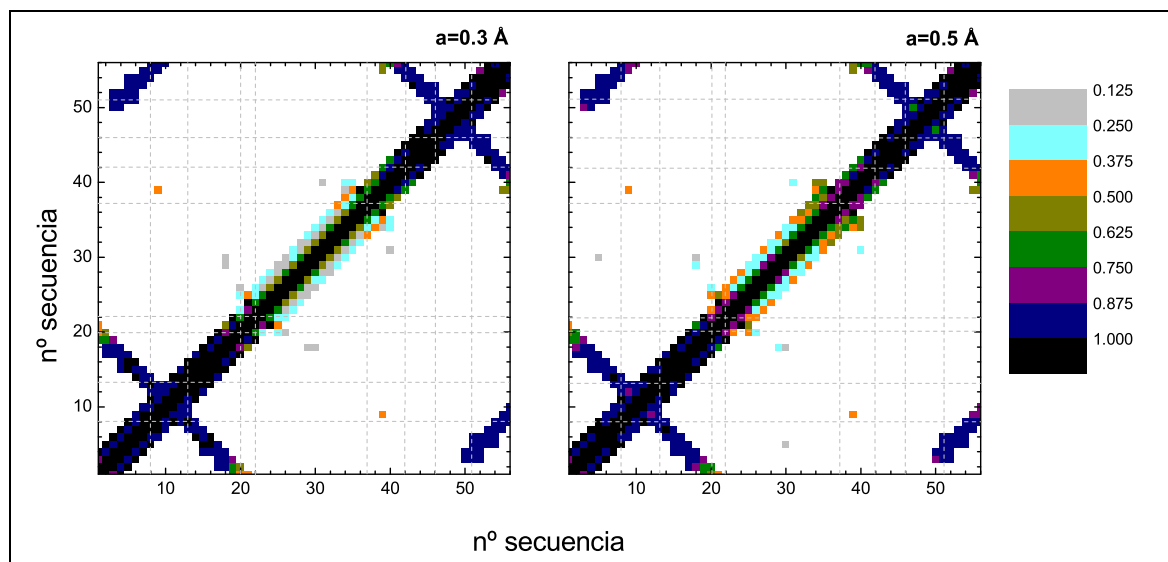


Figura 3.4: Estructura del intermedio de plegamiento que aparece con $d_{max} = 7.5 \text{ \AA}$. A la izquierda aparece la frecuencia de los contactos nativos para $a = 0.3 \text{ \AA}$ y a la derecha para $a = 0.5 \text{ \AA}$. La escala de colores indica esta frecuencia, calculada para la temperatura del mínimo de C_v^* entre los dos máximos. Esa temperatura es $T^* = 0.47$ en el caso de $a = 0.3 \text{ \AA}$ y $T^* = 0.52$ en el caso de $a = 0.5 \text{ \AA}$.

mina está esencialmente formada. Al disminuir la temperatura, la hélice se forma y la proteína adquiere el estado nativo. De hecho, en todas las simulaciones para todas las combinaciones de parámetros llegamos al estado nativo, validando así nuestro muestreo conformacional y el método de simulación.

Aunque algunos estudios sobre el mecanismo de plegamiento han postulado que una de las hebras β se forma antes que otros elementos estructurales, no se está hablando aquí de intermedios dinámicos de plegamiento. Los resultados de capacidad calorífica para $d_{max} = 7.5 \text{ \AA}$ indican la presencia de un intermedio termodinámico, que no corresponde a las observaciones experimentales. A mayores valores de a , las curvas de capacidad calorífica dejan de tener dos picos y pasan a presentar un pico y un hombro. Esto quiere decir que el intermedio termodinámico de plegamiento se hace menos nítido al aumentar la anchura del potencial.

La explicación para el diferente comportamiento que se observa con distintos valores de d_{max} se encuentra en los mapas de contactos a que dan lugar. Comparando estos

mapas en la Figura 3.2, se observan claramente las diferencias entre ellos, si bien ambos recogen, como ya se ha dicho, las características generales de la estructura. Las diferencias más importantes entre los dos mapas se encuentra en la zona de contacto entre la hélice y las hebras de la lámina β . Esta zona se encuentra ampliamente poblada en el caso de $d_{max} = 10$ Å mientras que para $d_{max} = 7.5$ Å apenas hay contactos en la misma. Debido a esto, para el valor más pequeño de d_{max} la hélice se encuentra desestabilizada frente a la lámina, ya que los únicos contactos que estabilizan la hélice son contactos locales. Dada la mayor estabilidad de la lámina, en un proceso de calentamiento ésta se deshace a temperatura mayor que a la que se deshace la hélice.

A la vista de los resultados, queda claro que es de gran importancia encontrar el valor de d_{max} más apropiado. Este valor variará de una estructura a otra en función de las características del mapa de contactos que representa su estado nativo. Por lo tanto, para poder describir de manera correcta la termodinámica del plegamiento, es necesario disponer de un mecanismo para determinar cuál es el valor más apropiado. Para ello, hay que profundizar en la estructura de la proteína a un nivel de detalle mayor que el que proporciona la representación de C_α que estamos utilizando. Una forma detallada y más realista de definir contactos entre aminoácidos en la estructura nativa es hacerlo en función del empaquetamiento de los átomos de las cadenas laterales, que no están incluidos en la representación de la proteína que estamos usando. A partir de los datos estructurales recogidos en el PDB se pueden determinar cuáles son estos contactos. En la Figura 3.5 mostramos los contactos que se obtienen teniendo en consideración todos los átomos pesados (no hidrógenos) de la proteína. También en este caso se establece una distancia de corte para determinar si dos aminoácidos están en contacto. Sin embargo, esta distancia de corte no se elige de manera tan arbitraria como en el caso de distancias entre C_α . Hemos elegido como distancia de corte un valor ligeramente superior al promedio del doble de los radios de van der Waals de los átomos pesados (es decir, los que forman parte de las proteínas: oxígeno, nitrógeno, carbono y azufre)¹³⁹. De esta forma,

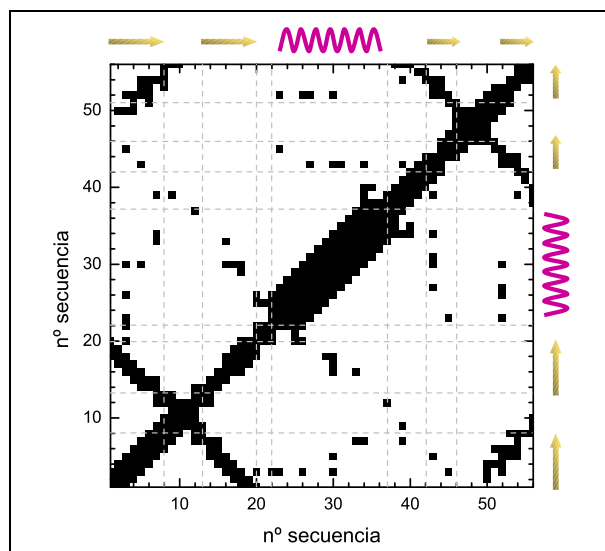


Figura 3.5: Mapa de contactos resultante de medir distancias entre los átomos pesados de los aminoácidos. La distancia de corte es 4.5 \AA .

si dos átomos pesados de residuos diferentes se encuentran a una distancia menor que este valor (4.5 \AA), se puede decir que esos átomos están en contacto en el estado nativo. Es decir, ese par de aminoácidos forma un contacto en el estado nativo, que se incluirá en el potencial de nuestro modelo.

El mapa de contactos de 2GB1 que resulta de tener en cuenta las distancias entre átomos pesados es el que aparece en la Figura 3.5. Se observa que este mapa incluye contactos entre la lámina β y la hélice. La presencia de contactos tanto locales como de largo alcance entre todos los elementos es, posiblemente, lo que hace que esta proteína se pliegue de manera cooperativa, sin intermedios termodinámicos. El intermedio que se detecta utilizando $d_{max} = 7.5 \text{ \AA}$ se debe a una estabilización relativa de la lámina sobre la hélice que no se corresponde con las observaciones experimentales. Este mapa está incorrectamente definido, ya que le faltan bastantes contactos entre elementos de estructura secundaria de los que aparecen en el mapa de resolución atómica. Sin embargo, el mapa que se obtiene con $d_{max} = 10 \text{ \AA}$ tiene todos esos contactos. De hecho, tiene muchos más. Se podría refinar más el mapa utilizando distancias de corte entre C_α más pequeñas que dieran lugar a mapas más parecidos al de la Figura 3.5. Dado que

queda clara la importancia del valor de d_{max} , hemos considerado innecesario incluir aquí ese estudio. Los resultados obtenidos demuestran que es necesario considerar todos los contactos reales en el esquema de interacciones simplificado para obtener resultados comparables con las observaciones experimentales. Por lo tanto, para hacer un estudio del plegamiento de proteínas utilizando un potencial basado en topología, es necesario determinar el valor de d_{max} entre C_α más apropiado para cada estructura.

El mapa con $d_{max} = 10 \text{ \AA}$ incluye todos los contactos relevantes del estado nativo de 2GB1, si bien es cierto que, como se ha dicho, es una distancia demasiado larga. Dado que los resultados obtenidos con este valor son comparables a las observaciones experimentales, en la siguiente sección utilizamos este valor para continuar con el análisis.

3.3. Anchura del potencial.

En la Figura 3.3 se puede ver cómo el valor de a afecta a la anchura del pico de C_v^* y a la temperatura a la que éste aparece. A medida que el parámetro a crece, la transición transcurre a lo largo de un intervalo de temperaturas mayor y, además, el máximo de capacidad calorífica aparece a mayor temperatura. Independientemente del valor de a , en el estado nativo todas las contribuciones a la energía están en el mínimo. Esto quiere decir que el estado nativo tiene la misma energía, que viene determinada por el número de contactos que se han definido, para cualquier valor de a . Si la energía del estado nativo fuera el único factor que determina su estabilidad, todas las curvas de capacidad calorífica deberían presentar un máximo a la misma temperatura. Sin embargo, vemos que esto no sucede así. Por lo tanto, otras contribuciones, diferentes a la energía del estado nativo, contribuyen también a su estabilidad.

Para realizar un análisis más profundo de las propiedades termodinámicas del sistema, hemos calculado el histograma de energía a distintas temperaturas y el perfil de energía libre a la temperatura de transición utilizando el método WHAM. En la Figura 3.6 se muestran los resultados de la distribución de energía para las distintas anchuras

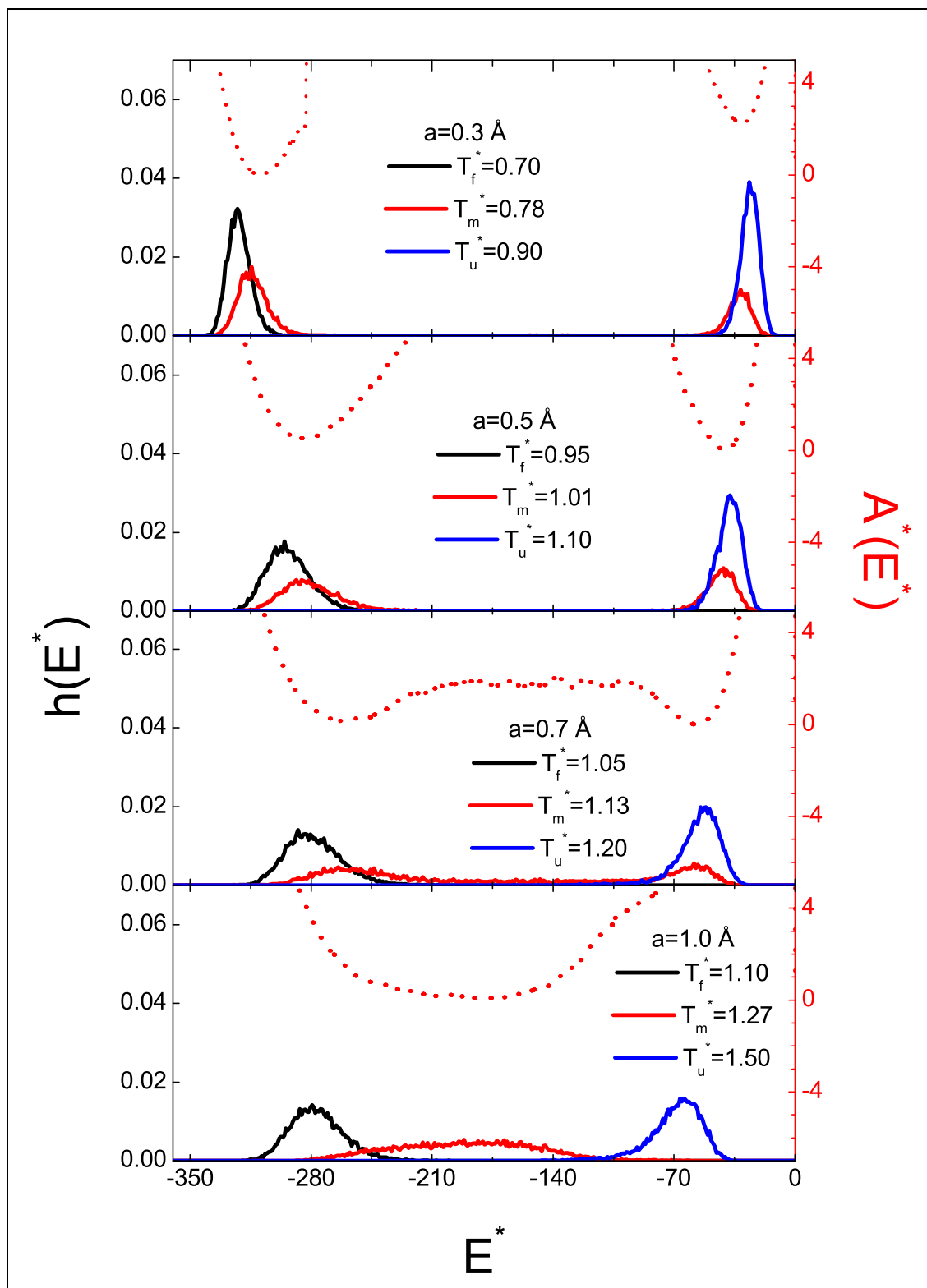


Figura 3.6: Histogramas de energía para temperaturas $T_f^* < T_m^*$, T_m^* y $T_u^* > T_m^*$ (línea continua) y perfil de energía libre a T_m^* (línea punteada). La escala de los histogramas se encuentra en el eje de la izquierda (negro) y la de la energía libre en el eje derecho (rojo). Cada panel corresponde a un valor de a diferente.

de potencial a tres temperaturas: a la temperatura de transición o plegamiento (T_m^*), que corresponde al máximo de capacidad calorífica, y a una temperatura ligeramente superior y otra ligeramente inferior a T_m^* en cada caso. Para todos los valores de a aparece un único estado a temperaturas mayores e inferiores a T_m^* , definido por un pico en el histograma de energía. A $T_f^* < T_m^*$ el pico en el histograma aparece a energías más negativas, como corresponde al estado plegado. En el caso de temperatura alta ($T_u^* > T_m^*$) el pico en el histograma, correspondiente al estado desnaturalizado, aparece a valores altos de energía. Dado que en el potencial no se han incluido contribuciones positivas, el estado desnaturalizado tiene todavía energía negativa, debido a la presencia de algunos contactos nativos que se forman esporádicamente, o a la presencia de cierta estructura nativa local. Por último, el histograma de energía correspondiente a T_m^* muestra diferencias de unas anchuras de potencial a otras. Para potenciales estrechos (valores pequeños de a) este histograma presenta una distribución bimodal, con dos picos. Sin embargo, para valor más alto de a (1.0 \AA) el histograma muestra un único pico a T_m^* que se va desplazando en la escala de energía cuando la temperatura aumenta o disminuye.

La distribución bimodal de energía es característica de un tránsito de dos estados. Los dos picos que aparecen corresponden al estado plegado y al desnaturalizado en equilibrio, habiendo población prácticamente nula para valores de energía intermedios entre esos dos estados. Un tránsito de dos estados es consecuencia de la cooperatividad de las interacciones. Para las anchuras de potencial $a = 0.3, 0.5$ y 0.7 \AA , la transición de plegamiento aparece descrita como una transición termodinámica de primer orden. Como podemos ver en los perfiles de energía libre en los paneles superiores de la Figura 3.6, el estado plegado y el desnaturalizado están separados por una barrera de energía libre que da lugar a una transición cooperativa de tipo “todo o nada”, considerada característica del plegamiento de proteínas. Sin embargo, para $a = 1.0 \text{ \AA}$, la distribución de energía a T_m^* presenta un único pico a valores de energía intermedios entre el estado plegado y el desnaturalizado. Esta distribución corresponde a un plegamiento sin barrera¹² (o de

tipo *downhill*), en el cual se adquiere la estructura del estado nativo de manera gradual a medida que se varían las condiciones (en este caso, de temperatura), lo cual queda reflejado en el perfil de energía libre a T_m^* con un único mínimo. Un plegamiento de este tipo no corresponde a la transición experimental que tiene lugar en el plegamiento de la proteína que estamos estudiando. El valor de $a = 1.0 \text{ \AA}$, en principio, implica una desviación muy pequeña respecto del estado nativo, teniendo en cuenta la resolución del modelo. Sin embargo, es suficientemente grande como para dar resultados radicalmente distintos y, en este caso, incorrectos.

Por otra parte, tampoco es conveniente utilizar valores muy pequeños de a . En la Figura 3.6 vemos que, con los potenciales más estrechos ($a = 0.3$ y 0.5 \AA) la barrera de energía libre que aparece es enormemente elevada, lo cual hace que sea muy difícil de superar. En nuestro caso, implica utilizar un número de temperaturas muy elevado en el esquema de *parallel tempering*. Además, si comparamos los resultados de capacidad calorífica de las simulaciones con los datos experimentales de la misma proteína¹³¹, como hacemos en la Figura 3.7, vemos que las diferencias son notables. Para poder comparar las amplitudes de los picos de capacidad calorífica, es necesario normalizar las curvas de manera que estén referidas a una misma temperatura y tengan todas la misma altura. La amplitud del pico nos da una idea de lo abrupto de la transición. Normalizando respecto a la temperatura de transición y al máximo de cada curva, se puede ver que las transiciones que tienen lugar con valores de $a = 0.3$ y 0.5 \AA son muy estrechas comparadas con la transición experimental. Sin embargo, $a = 0.7 \text{ \AA}$ da resultados más comparables con los experimentales. Estas comparaciones deben hacerse con cuidado, ya que estamos comparando valores de C_p con valores de C_v . Pero a la vista de los resultados, podemos decir que $a = 0.7 \text{ \AA}$ es un valor apropiado para describir la transición de 2GB1 y que los potenciales mas estrechos dan lugar a transiciones demasiado abruptas. Además, para el potencial más ancho la transición ocurre en un intervalo de temperaturas de aproximadamente un 10 % del valor de T_m^* . En una escala real de temperatura, esto implicaría

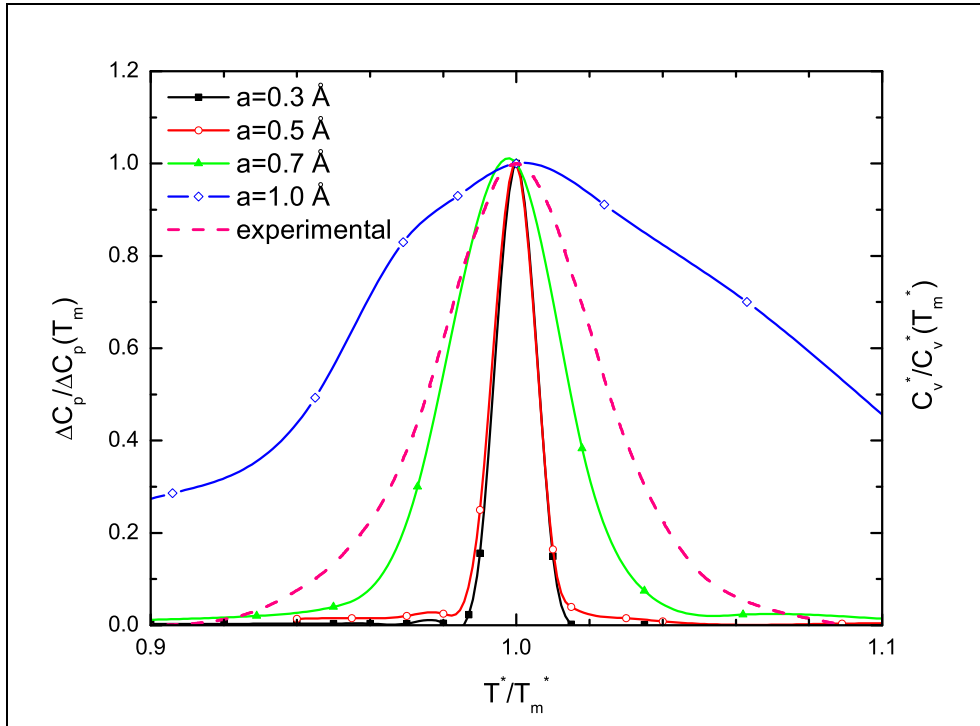


Figura 3.7: Comparación de las curvas de capacidad calorífica obtenidas con los distintos valores de a con la curva experimental. Las curvas están normalizadas con respecto al valor máximo de C_v^* o ΔC_p y a la temperatura de transición. La curva de capacidad calorífica experimental está medida a presión constante (eje de la izquierda) mientras que los resultados de las simulaciones se obtienen a volumen constante (eje de la derecha).

un intervalo de 30 K, que es un valor razonable teniendo en cuenta las condiciones experimentales reales en que tiene lugar el plegamiento. Esto valida el modelo de interacción propuesto en la forma de la ecuación 2.1.

Lo que queda claro a partir de los resultados presentados aquí es que el tipo de transición está enormemente influido por la anchura del potencial de atracción. Es importante tener en cuenta que este parámetro no depende de los contactos presentes en la proteína. Por lo tanto, podemos pensar que el mismo valor de a que es apropiado para una proteína lo es para otra, lo cual ha sido comprobado en otros trabajos realizados por nuestro grupo de investigación¹⁴⁰ y en resultados posteriores de esta Tesis.

Por otra parte, hemos visto que el valor de a tiene, además, una gran influencia en la estabilidad de la proteína, dado que se produce una variación en el valor obtenido de

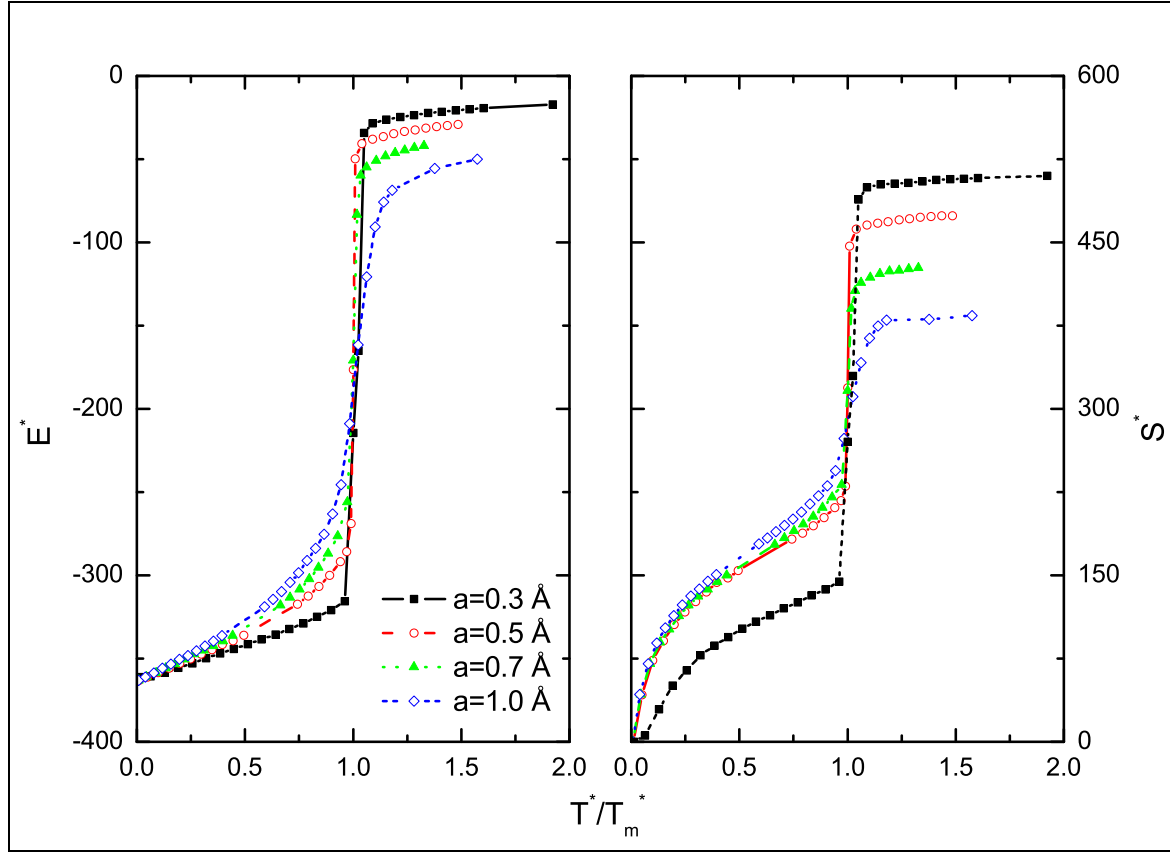


Figura 3.8: Variación de la energía (izquierda) y la entropía (derecha) del sistema con la temperatura. Las curvas están normalizadas con respecto a T_m^* .

T_m^* . *A priori* esto es sorprendente, ya que en un potencial basado en la topología como el que se utiliza aquí, el estado nativo está estabilizado por los mismos contactos independientemente de la anchura de potencial. Para poder profundizar en las propiedades de la transición que explican estas observaciones, en la Figura 3.8 representamos la variación de la energía y la entropía medias del sistema a cada temperatura calculadas con el método WHAM (ecuación (2.14)). Para poder comparar estas curvas, también han sido normalizadas con respecto al valor de T_m^* de cada una de ellas. También en esta figura se puede observar la diferente amplitud que tiene la transición dependiendo de la anchura del potencial. Pero, además, en esta figura es destacable que tanto la energía como la entropía del estado desnaturalizado se ven enormemente afectadas por el valor de a . El estado plegado presenta prácticamente las mismas propiedades en todos los casos,

excepto en la entropía para el caso de $a = 0.3 \text{ \AA}$ que, al ser un potencial muy estrecho, permite muy pocas fluctuaciones a temperaturas bajas, desviándose del comportamiento de los otros valores de a . Vemos que, efectivamente, a baja temperatura, el estado nativo tiene la misma energía en todos los casos. Por el contrario, el estado desnaturalizado presenta valores distintos de energía para cada valor de a y esta diferencia es aun mayor en los valores de entropía. Dado que el potencial que se está usando está basado en la topología del estado nativo, sin ningún tipo de información sobre la secuencia, sería de esperar que el estado desnaturalizado fuera el estado peor definido y menos afectado por las variaciones del potencial. Sin embargo, la influencia sobre las propiedades de este estado son claras y por lo tanto deben ser tenidas en cuenta a la hora de utilizar este tipo de potenciales.

En el estado nativo la energía total es $E = N_{con} \times u_{min}$, donde N_{con} es el número de contactos presentes en el estado nativo y u_{min} es la energía de cada contacto en el mínimo (en este caso $u_{min}^* = -1$). Independientemente de la anchura del potencial, el estado nativo tiene una energía determinada por el número de contactos. En un potencial más ancho, las fluctuaciones respecto al estado nativo están más favorecidas que en uno estrecho, ya que una pequeña variación en la distancia produce un aumento de energía mucho mayor en un potencial estrecho que en uno ancho (ver Figura 3.1). Cuanto más favorecidas están estas fluctuaciones respecto al estado nativo, mayores son la energía y entropía medias del sistema. Por este motivo, el estado plegado tiene energía y entropía ligeramente mayor para potenciales más anchos. Por otra parte, como ya hemos dicho, esta variación de energía y entropía de unos potenciales a otros es, en general, muy pequeña en el estado plegado, pero muy notable en el estado desnaturalizado. Un potencial más ancho asigna energía $u_{ij}^* < 0$ a contactos esporádicos para los que un potencial estrecho asigna energía $u_{ij}^* = 0$. En la Figura 3.1 se puede ver que un par de aminoácidos a una distancia d_{kl} , relativamente cercana a la nativa, tiene menor energía cuanto mayor es la anchura del potencial. Para el potencial más estrecho, este contacto

Tabla 3.1: Propiedades termodinámicas de la transición en unidades reducidas para las distintas anchuras del potencial. $d_{max} = 10 \text{ \AA}$.

| $a / \text{ \AA}$ | $T_m^*{}^a$ | ΔE_m^* | ΔS_m^* | $T_m^*{}^b$ |
|-------------------|-------------|----------------|----------------|-------------|
| 0.3 | 0.78 | 279 | 347 | 0.80 |
| 0.5 | 1.01 | 235 | 233 | 1.01 |
| 0.7 | 1.13 | 155 | 137 | 1.13 |
| 1.0 | 1.27 | ... | ... | ... |

^aCalculada a partir del máximo de capacidad calorífica.

^bCalculada a partir de la relación $T_m^* = \Delta E_m^* / \Delta S_m^*$

tiene una energía $u_{kl} = 0$. A altas temperaturas, donde la proteína no tiene estructura definida y los contactos nativos son esporádicos, el estado desnaturalizado tiene mayor energía residual (en valor absoluto) cuanto mayor es la anchura del potencial, como se puede ver en la Figura 3.8. Desde el punto de vista de una superficie de energía en forma de embudo, como corresponde a un potencial basado en la topología, estados de menor energía tienen asociada una menor entropía y esto, como muestra la Figura 3.8, coincide con los resultados de nuestras simulaciones.

Para poder explicar la diferente estabilidad de la proteína con los distintos tipos de potencial, debemos recurrir a los valores de ΔE^* y ΔS^* entre los estados que aparecen a T_m^* , es decir, el estado plegado y el desnaturalizado. Estas diferencias son las que determinan la estabilidad de la proteína. Tomando como T_m^* la temperatura a la cual aparece un máximo de capacidad calorífica, calculamos ΔE^* y ΔS^* a esa temperatura. Estos valores, ΔE_m^* y ΔS_m^* , aparecen en la Tabla 3.1. T_m^* aumenta, como ya hemos comentado, a medida que el potencial se hace más ancho. Según la ecuación (3.2) de la energía libre:

$$\Delta A^* = \Delta E^* - T^* \Delta S^* \quad (3.2)$$

y sabiendo que a $T^* = T_m^*$, $\Delta A^* = 0$, la relación entre la temperatura de transición y

las propiedades termodinámicas del tránsito es: $T_m^* = \Delta E_m^* / \Delta S_m^*$. Una disminución en el valor de ΔE_m^* implica un valor menor de T_m^* . Por el contrario, una disminución en el valor de ΔS_m^* provoca un aumento en el valor de T_m^* . A medida que aumenta la anchura de potencial, tanto ΔE_m^* como ΔS_m^* disminuyen. Los efectos de estas dos variaciones son contrarios y podrían hacer que T_m^* permaneciera aproximadamente constante. Sin embargo, como vemos en la Tabla 3.1, las diferencias en ΔE_m^* y ΔS_m^* no se compensan. Aunque ambos valores decrecen cuando aumenta el valor de a , la variación en entropía de unos potenciales a otros es más acentuada que la variación de energía. Por ese motivo, el valor de T_m^* aumenta con el parámetro a . Si recalculamos T_m^* a partir de ΔE_m^* y ΔS_m^* (ver Tabla 3.1), vemos que los valores de T_m^* así obtenidos coinciden con bastante exactitud con la del máximo de capacidad calorífica, validando así los resultados obtenidos. La única diferencia que encontramos ocurre para $a = 0.3 \text{ \AA}$. Para este potencial, la transición es tan abrupta que un análisis detallado se hace mucho más difícil. Esto deja claro una vez más que un potencial demasiado estrecho resulta poco adecuado para el estudio de la transición del plegamiento con un potencial basado en la topología.

Hemos visto, por tanto, que la anchura de potencial, debido a su influencia sobre el estado desnaturalizado, determina las propiedades de la transición de plegamiento. Además, para poder sacar conclusiones sobre la estabilidad de proteínas utilizando potenciales de este tipo, no es suficiente con tener en consideración la energía del estado nativo. Es necesario realizar un análisis detallado de las propiedades del estado desnaturalizado, que es un estado frecuentemente olvidado cuando se están utilizando estos potenciales basados en la topología del estado nativo.

3.4. Mapas de contactos con resolución atómica.

En la Sección 3.2, hemos utilizado el mapa de contactos “real”, basado en distancias entre átomos pesados, como referencia para determinar cuál es el valor más apropiado de d_{max} entre C_α , que dependerá de la estructura de la proteína que se quiere estudiar.

Sin embargo, si utilizamos directamente el mapa de contactos que tiene en cuenta todos los átomos pesados de la proteína, no será necesario determinar el valor de d_{max} más apropiado en cada caso. A partir de este mapa se determina qué pares de aminoácidos están formando un contacto en el estado nativo. Una vez hecho esto, se consideran las distancias entre los átomos C_α de esos pares de aminoácidos como distancia d_{ij}^{nat} en las ecuaciones (2.1) y (3.1). Las interacciones se siguen calculando entre los C_α para aquellos pares que forman un contacto en el estado nativo. Lo que cambia aquí es la manera de determinar cuáles son esos pares. Para ello se toma la proteína con todos los átomos que la constituyen y se mide la distancia más corta entre átomos pesados de aminoácidos diferentes. Si esa distancia es igual o menor que 4.5 Å que, como se ha dicho, es una media de la suma de los radios de van der Waals de los átomos pesados que se encuentran en las proteínas, se considera que esos aminoácidos forman un contacto en el estado nativo. Para 2GB1, el mapa que se obtiene así es el que ya hemos comentado de la Figura 3.5. Para los pares de aminoácidos marcados en ese mapa como contactos, el valor de $u_{ij} \neq 0$, mientras que para el resto $u_{ij} = 0$.

Por otra parte, en la sección anterior se ha llegado a la conclusión de que, en relación a los datos experimentales, el valor $a = 0.7$ Å es el que ofrece mejores resultados. Con este valor se evitan, además, ciertos problemas de tipo metodológico. Un potencial demasiado estrecho da lugar a transiciones muy abruptas que dificultan el estudio de las mismas y que, además, no corresponden a las observaciones experimentales.

En esta parte del estudio hemos comprobado que el mapa de contactos con resolución atómica, combinado con un potencial de interacción entre C_α de anchura $a = 0.7$ Å, da buenos resultados. En la Figura 3.9a se representa la variación de la capacidad calorífica con la temperatura. Para poder comparar los resultados de estas simulaciones con los resultados experimentales, hemos normalizado las curvas de la manera ya explicada anteriormente (ver Figura 3.7). Podemos ver que la amplitud del pico obtenida con este modelo es comparable con los resultados experimentales. En este caso, la temperatura

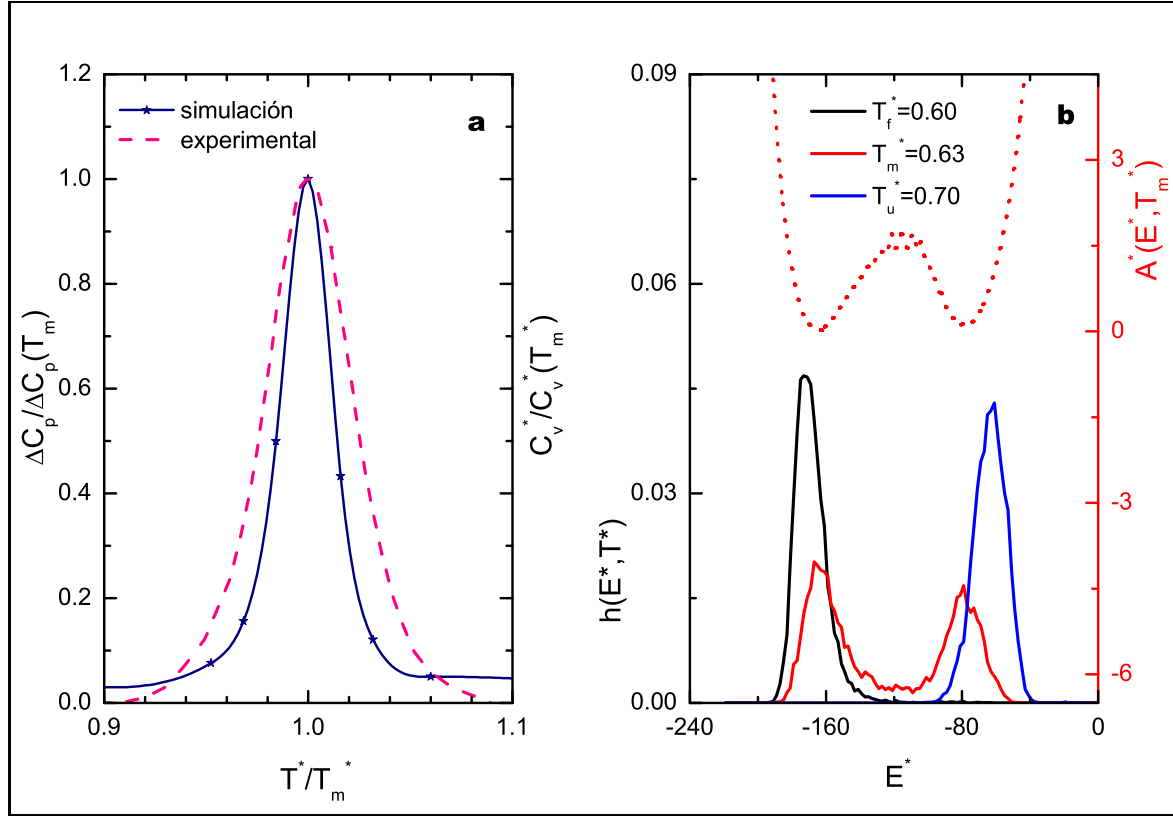


Figura 3.9: **a)** Comparación de las curvas de capacidad calorífica experimental y la obtenida a partir de los resultados de la simulación. Estas curvas están normalizadas con respecto al valor máximo de C_v^* o ΔC_p y a la temperatura de transición. La curva de capacidad calorífica experimental está medida a presión constante (eje de la izquierda) mientras que los resultados de las simulaciones se obtienen a volumen constante (eje de la derecha). **b)** Histograma de energía de las simulaciones a las temperaturas $T_f^* < T_m^*$, T_m^* y $T_u^* > T_m^*$ (líneas continuas) y perfil de energía libre a T_m^* (línea punteada). La escala de los histogramas se encuentra en el eje de la izquierda (negro) y la de la energía libre en el eje derecho (rojo).

de transición es $T_m^* = 0.63$. Este valor es inferior a la T_m^* que se obtiene con el mapa de $d_{max} = 10 \text{ \AA}$ entre C_α , lo cual está relacionado con el mayor número de contactos que se definen para el estado nativo con el mapa de contactos obtenido a partir de distancias entre C_α .

Para poder estudiar en detalle las propiedades de la transición según el mapa de contactos definido, hemos calculado los histogramas de energía a tres temperaturas diferentes y el perfil de energía libre a T_m^* . En la Figura 3.9b vemos que la distribución de energía a T_m^* es bimodal, como corresponde a una transición de dos estados y que,

además, el perfil de energía libre a esa temperatura presenta una barrera característica de este tipo de transiciones.

A la vista de estos resultados podemos decir que una anchura de potencial de $a = 0.7 \text{ \AA}$, combinado con el mapa de contactos definido por los átomos pesados con $d_{max} = 4.5 \text{ \AA}$, da buenos resultados a la hora de estudiar el plegamiento de proteínas. Hemos obtenido así un potencial basado en la topología del estado nativo, o de tipo $G\bar{o}$, que pensamos puede ser generalizable a todas las estructuras. Podemos entonces pasar a utilizarlo para ver cómo la topología del estado nativo determina las propiedades de las transiciones de plegamiento y agregación de proteínas, como haremos en los siguientes capítulos. Antes de eso, sin embargo, nos queda todavía por analizar un aspecto metodológico relevante.

3.5. Estructura de una proteína determinada por diferentes técnicas.

Uno de los problemas que presenta un potencial basado en la topología del estado nativo es precisamente cómo está determinada esa topología. Las coordenadas de los átomos en el estado nativo se obtienen de los datos del Protein Data Bank (PDB)²⁶. El PDB es una base de datos donde se pueden encontrar las coordenadas de todas las estructuras que han sido resueltas. Estas estructuras son determinadas, principalmente, a partir de difracción de Rayos X o de experimentos de Resonancia Magnética Nuclear (RMN). Además de los datos estructurales, en esta base de datos se puede encontrar información general sobre las proteínas recogidas y sobre el método de determinación estructural utilizado.

Para el estudio que hemos realizado hasta aquí hemos utilizado la estructura cuyo código PDB es 2GB1, que corresponde a la determinación mediante RMN de la estructura del dominio B1 de la proteína G de *Streptococcus*. Además, y bajo el código 1PGB,

en el PDB aparece recogida la estructura de esta proteína determinada por difracción de Rayos X. Tratándose de la misma proteína, es de esperar que la estructura determinada por ambos métodos sea idéntica o, por lo menos, muy parecida. Efectivamente, el valor de *rmsd* (*root mean square deviation*, ver Capítulo 2) entre las dos estructuras es de 1.18 Å, que es un valor relativamente pequeño. De hecho, si comparamos el mapa de contactos de cada una de las estructuras, vemos que son muy similares. En la parte superior de la Figura 3.10 podemos ver el mapa correspondiente a 1PGB y a 2GB1, ambos calculados a partir de distancias entre átomos pesados, y comprobamos que ambos mapas se parecen mucho. Sin embargo, en la parte inferior de la figura, donde se representan los contactos que aparecen en un mapa y no en el otro, vemos que existen algunas diferencias entre ellos. ¿Cómo afecta esto al estudio con potenciales basados en la topología?

Para comprobar cómo de significativas son las diferencias entre los dos mapas, hemos realizado varias simulaciones con cada uno de ellos y hemos calculado las diferentes propiedades termodinámicas de las transiciones a que dan lugar. En la Figura 3.11a aparece representada la capacidad calorífica normalizada. Se puede ver que ambas curvas son muy similares, aunque el pico de la curva de C_v^* que corresponde a 1PGB es un poco más ancho que el de la curva de 2GB1, lo cual es indicativo de una transición algo más abrupta para 2GB1 que para 1PGB. Además, las temperaturas de transición de 2GB1 y 1PGB también son ligeramente distintas, siendo 0.63 y 0.65, respectivamente. Los perfiles de energía libre y los histogramas de energía a T_m^* de la Figura 3.11b corresponden en ambos casos a una transición de primer orden. Sin embargo, la barrera de energía en el caso de 2GB1 es mayor que en el caso de 1PGB, lo cual indica que la transición que tiene lugar para 2GB1 es más abrupta que en el caso de 1PGB, como sugieren las curvas de capacidad calorífica. Por otra parte, el pico de mayor energía en el histograma o en el perfil de energía libre indica que el estado desnaturalizado que se describe con nuestro modelo en ambos casos tiene las mismas características. Podemos decir entonces que

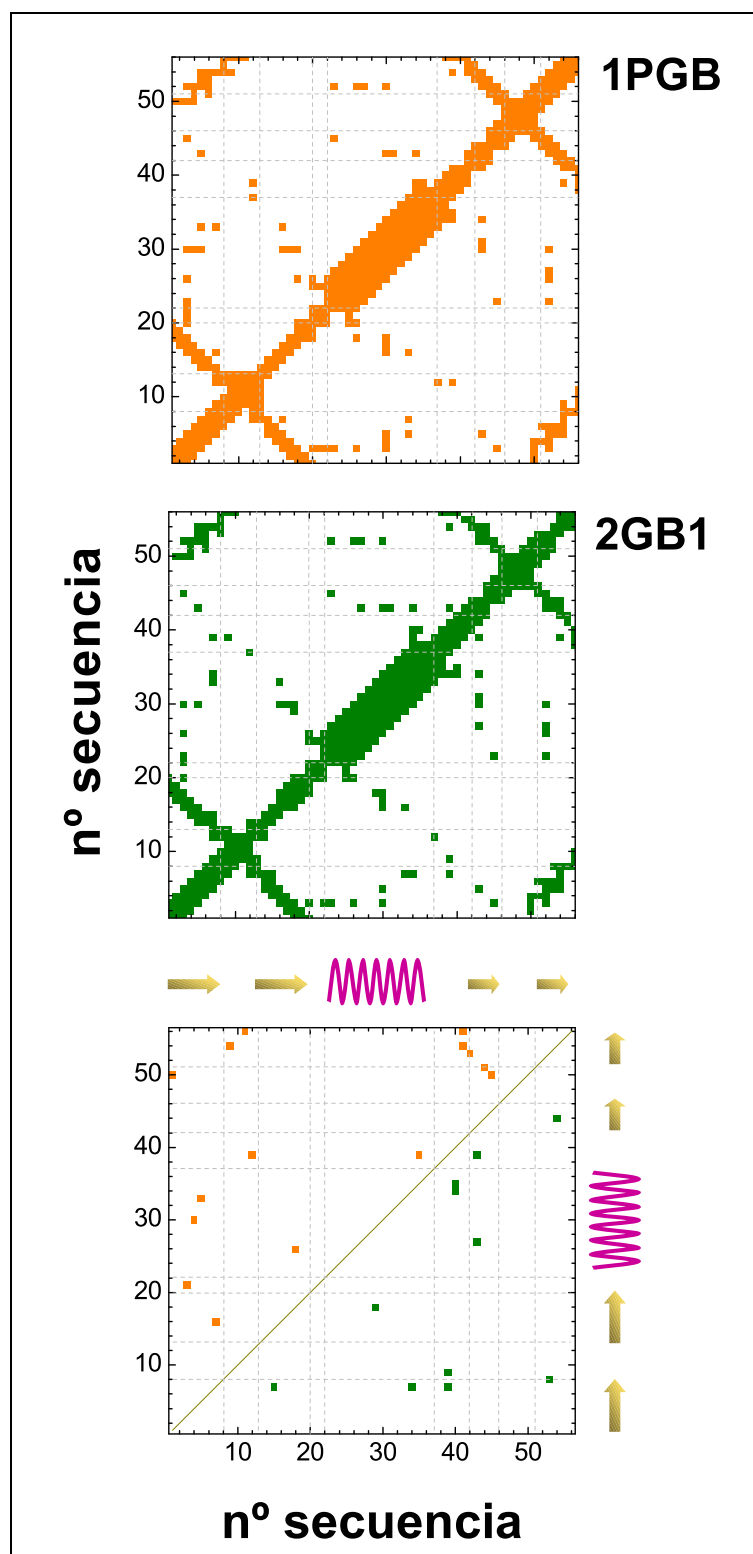


Figura 3.10: Mapa de contactos del dominio B1 de la proteína G de *Streptococcus* a partir de rayos X (1PGB) y de RMN (2GB1). El mapa de abajo representa, en el triángulo superior izquierdo, los contactos que hay en 1PGB que no están en 2GB1 y en el triángulo inferior derecho, los contactos que tiene 2GB1 que no aparecen en 1PGB.

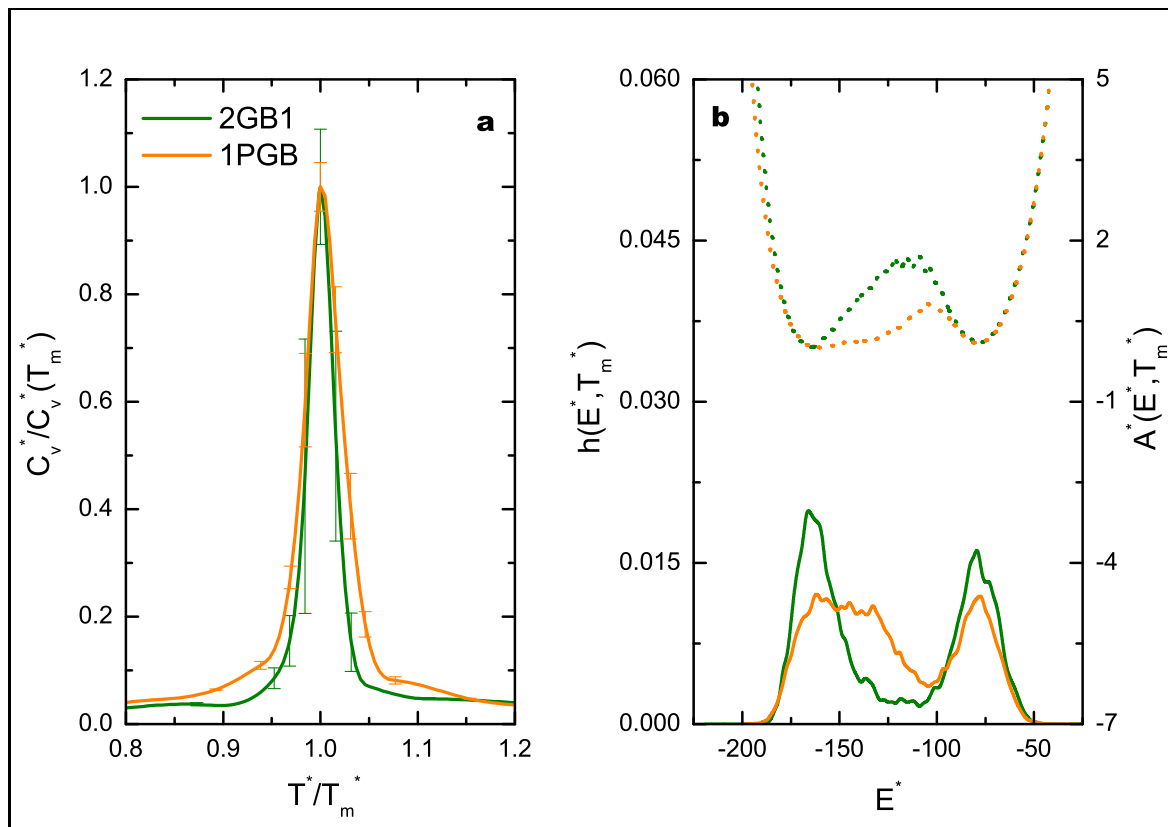


Figura 3.11: Propiedades termodinámicas de la transición de plegamiento descrita con los mapas deducidos a partir de 2GB1 (verde) y 1PGB (naranja): A la izquierda, curvas de capacidad calorífica a volumen constante normalizadas y a la derecha histograma de energía (línea continua) y perfil de energía libre (línea punteada) a la temperatura de transición T_m^* .

este estado se ve menos afectado por las diferencias en los mapas de contactos definidos. Sin embargo, estas diferencias tienen una gran influencia en las propiedades del estado plegado que se define a T_m^* . Centrándonos en la curva de energía libre, vemos que el mínimo correspondiente al estado de baja energía es más ancho en el caso de 1PGB, es decir, este estado está peor definido, en términos de variabilidad conformacional, a la temperatura de transición en el caso de 1PGB que en el de 2GB1: para 1PGB el estado plegado tiene una mayor libertad conformacional a T_m^* . Para poder profundizar en cuáles son las diferencias estructurales que podrían dar lugar a la diferencia en el comportamiento termodinámico de estos dos modelos, nos fijamos en la parte inferior de la Figura 3.10, en la que representamos las diferencias entre los dos mapas. En el

triángulo superior izquierdo aparecen los contactos presentes en 1PGB que no están en 2GB1 y en el triángulo inferior derecho, lo contrario. El mapa de contactos de 1PGB tiene 15 contactos que 2GB1 no tiene y 2GB1 tiene 11 contactos que 1PGB no tiene. En total, 1PGB tiene 4 contactos más. Algunas de las diferencias que se observan son sutiles, como por ejemplo, el contacto entre los aminoácidos 7 y 16, que aparece en 1PGB pero no en 2GB1, es un contacto entre la hebra $\beta 1$ y la 2. Entre estos dos elementos de estructura secundaria, en 2GB1 también aparece un contacto (entre los aminoácidos 7 y 15). Podemos decir que son contactos equivalentes. Las diferencias mayores son aquellas que corresponden a contactos entre elementos de estructura secundaria en uno de los mapas que no tienen un equivalente en el otro. Algunas diferencias importantes se encuentran en la zona de contactos entre la hebra $\beta 3$ y el lazo de unión entre las hebras 3 y 4, donde 1PGB presenta contactos que 2GB1 no tiene. Por otra parte, 2GB1 presenta contactos entre la hebra $\beta 3$ y la hélice o el lazo entre la hélice y la hebra 3 que 1PGB no presenta.

La cuestión que se plantea es, ¿cuáles de esas diferencias entre los mapas son las que determinan que la transición sea de una manera u otra? Es decir, ¿son los contactos presentes en 1PGB que faltan en 2GB1 los que hacen disminuir la barrera de energía libre o, por el contrario, son los contactos no presentes en 1PGB, que sí están en 2GB1, los que dan lugar a esta diferencia en el comportamiento termodinámico? Para dar respuesta a estas preguntas, hemos planteado una serie de experimentos numéricos modificando artificialmente el mapa de contactos de 1PGB. De esta forma podemos definir que en el potencial del modelo dos aminoácidos estén en contacto o no, al margen de lo que indica el mapa, según nuestro interés. De esta manera hemos creado una especie de *mutantes* de 1PGB. Estos mutantes se han generado “apagando” o “encendiendo” contactos del mapa. Es decir, definiendo que no hay contacto entre dos residuos que sí forman uno en 1PGB (con lo cual deja de formar parte del esquema de interacción), o asignando como contacto de cara al cálculo de energía un par de aminoácidos que no están lo suficientemente cerca

en la estructura determinada por rayos X de acuerdo a la distancia de corte considerada. De esta forma, se han definido cuatro mapas de contactos diferentes que podemos ver esquematizados en la Figura 3.12. Las características de estos mapas son las siguientes:

- **Mapa 1:** 1PGB tiene 4 contactos más que 2GB1. Por lo tanto, la primera modificación que hemos hecho es partir del mapa de 1PGB y eliminar en el mismo 4 contactos que están presentes en 1PGB pero no en 2GB1. Como se puede ver en la Figura 3.10, 1PGB tiene más contactos que 2GB1 en las zonas de interacción de las hebras con los lazos. Por lo tanto, hemos “desactivado” cuatro contactos de esas zonas que no tengan un equivalente en 2GB1. Esos contactos son (1-50), (3-21), (11-56) y (42-53), marcados en azul en la Figura 3.12. El primero corresponde a la zona de contacto entre la hebra 1 y el lazo entre las hebras 3 y 4. Los aminoácidos 3 y 21 pertenecen a la hebra 1 y a la zona intermedia entre la hebra 2 y la hélice. Los dos últimos contactos se producen entre la hebra 4 y los lazos entre las hebras 1 y 2 y entre la hélice y la hebra 3, respectivamente.
- **Mapa 2:** Por otro lado, 2GB1 también tiene contactos que no presenta 1PGB, aun teniendo un menor número de contactos. Dichos contactos corresponden, principalmente, a la interacción entre la lámina β y la hélice α . En este caso hemos partido también del mapa de 1PGB, añadiendo al mismo cuatro de estos contactos existentes en 2GB1 (en gris en la Figura 3.12): (7-39), que se da entre la hebra 1 y el lazo entre la hebra 3 y la hélice; (34-40), entre la hélice y el lazo entre la hélice y la hebra 3; (39-43) que es un contacto entre este mismo lazo y la hebra 3; y (27-43), que se da entre la hélice y la hebra 3.
- **Mapa 3:** Este mapa es una combinación de los dos anteriores, por lo que mantiene el mismo número de contactos que 1PGB, aunque 4 de ellos son de diferente naturaleza.
- **Mapa 4:** Para reducir el número de contactos de 1PGB a un número igual al que

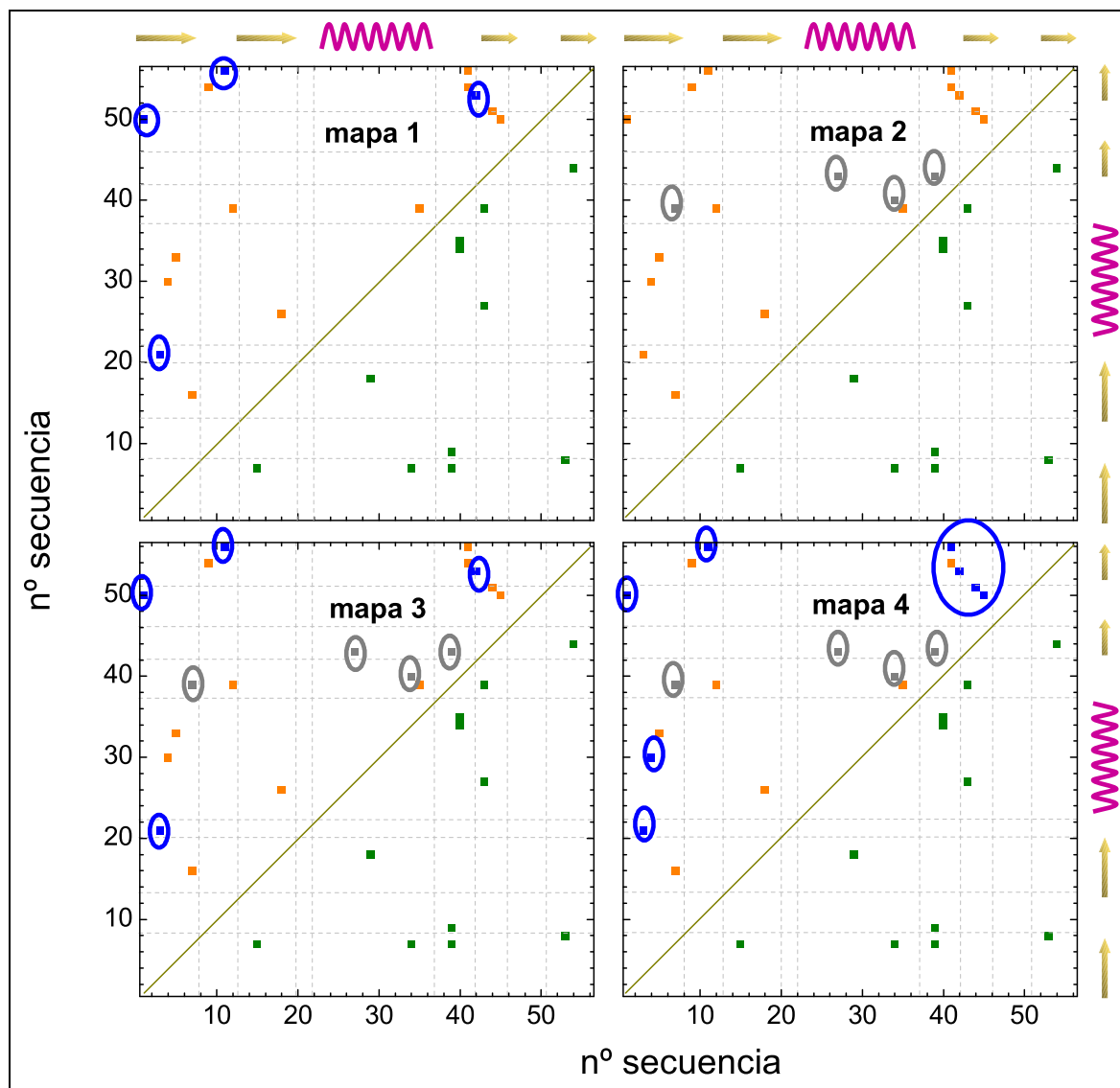


Figura 3.12: En cada mapa se representan: en verde, los contactos presentes en 2GB1 y que no aparecen en 1PGB; en naranja, los que aparecen en 1PGB y no en 2GB1; en azul, marcados con un círculo, los contactos que se “apagan” en 1PGB y en gris, también marcados con un círculo, los que se “encienden”.

Tabla 3.2: Propiedades termodinámicas de la transición en unidades reducidas para los distintos mapas diseñados. N_{lr} es el número de contactos de largo alcance presentes en el mapa de contactos.

| | N_{lr} | T_m^* | ΔE_m^* | ΔS_m^* |
|---------------|----------|---------|----------------|----------------|
| 1PGB | 112 | 0.65 | 111 | 171 |
| mapa 1 | 108 | 0.64 | 112 | 176 |
| mapa 2 | 116 | 0.66 | 119 | 179 |
| mapa 3 | 112 | 0.65 | 115 | 178 |
| mapa 4 | 108 | 0.63 | 113 | 179 |

presenta 2GB1, hemos partido del **mapa 3** y hemos desactivado cuatro contactos más de la zona principal donde 1PGB tiene contactos en exceso con respecto a 2GB1. Dichos contactos son (4-30), (41-56), (44-51) y (45-50). Apagando estos contactos que no tienen equivalente en 2GB1, tenemos el mismo número de contactos y los mapas se asemejan más. Los contactos que se eliminan corresponden a interacciones entre la hebra 1 y la hélice, la hebra 4 y el lazo entre la hélice y la hebra 3, y la hebra 3 y el lazo entre las hebras 3 y 4, zona a la que corresponden los dos últimos contactos.

Para evaluar las consecuencias en la termodinámica del tránsito de modificar el mapa de contactos, en la Tabla 3.2 indicamos los valores de T_m^* y de ΔE_m^* y ΔS_m^* de las transiciones a que dan lugar los distintos mapas de la Figura 3.12. Vemos que la temperatura de transición apenas se modifica de unos casos a otros. Sin embargo sí pueden apreciarse algunos cambios en la diferencia de energía y entropía entre los estados plegado y desnaturalizado. Para analizar estas variaciones, en la Figura 3.13 mostramos los perfiles de energía libre calculados a T_m^* para las transiciones definidas por estos mapas. Vemos que el mínimo que aparece a valores poco negativos de energía, correspondiente al estado desnaturalizado, no varía sustancialmente de unos casos a

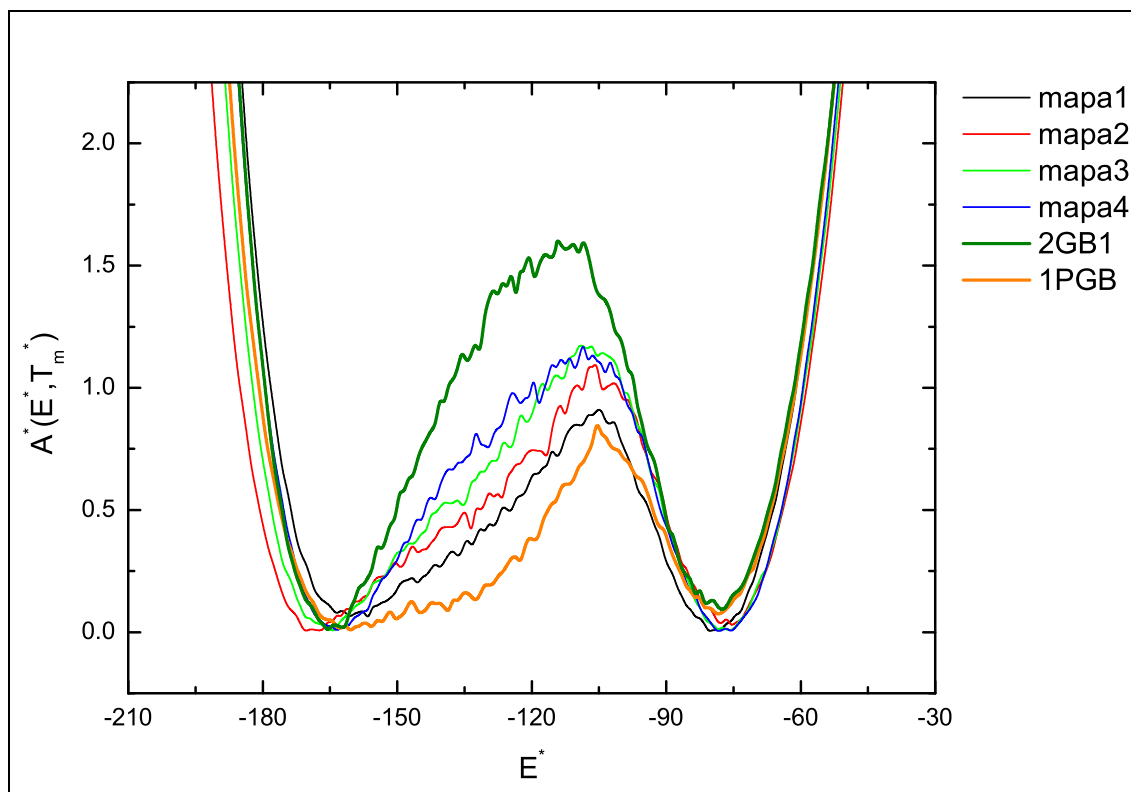


Figura 3.13: Perfil de energía libre de la transición termodinámica descrita por los distintos mapas de contactos de la Figura 3.12

otros. Sin embargo, el mínimo que corresponde a estados de baja energía varía de unos casos a otros. Es importante observar cómo afectan las mutaciones realizadas a las variaciones en ΔE_m^* y ΔS_m^* , aunque éstas no son lo suficientemente grandes como para dar lugar a una variación notable de la temperatura de transición.

- En el caso del **mapa 1**, en el que hemos eliminado contactos que hacen que la lámina β esté estabilizada con respecto a la hélice, se produce un ligero estrechamiento del mínimo que corresponde al estado de baja energía a esa temperatura. Como se puede ver en la Tabla 3.2 esto afecta, principalmente al valor de ΔS_m^* . Dado que el mínimo se estrecha, se produce una disminución en la entropía del estado plegado con respecto a 1PGB sin modificar, que da lugar al aumento en ΔS_m^* .
- Por otra parte, añadir contactos al mapa de manera que aumente la estabilidad

de la hélice α , como hacemos en el **mapa 2**, hace que el mínimo de energía libre a T_m^* que corresponde al estado plegado se estreche algo más que en el caso anterior. Este estrechamiento hace que disminuyan la energía y la entropía medias de este estado, aumentando tanto ΔE_m^* como ΔS_m^* .

- Cuando combinamos los **mapas 1 y 2** los efectos anteriores se compensan, dando lugar a valores de ΔE_m^* y ΔS_m^* intermedios entre los del **mapa 1** y los del **2**.
- Por último, eliminar más contactos de la zona que estabiliza la lámina β con respecto a la hélice α produce una disminución en ΔE_m^* (respecto al **mapa 3**), debido al menor número de contactos en el estado nativo, que combinado con el ligero aumento en ΔS_m^* da lugar a una pequeña disminución en T_m^* . En este caso, ΔE_m^* es muy similar al valor de esta propiedad para el **mapa 1**, sin embargo, la diferente distribución de los contactos provoca una mejor definición del estado de baja energía, que se comprueba con el estrechamiento del mínimo correspondiente a dicho estado. Debido a este hecho, ΔS_m^* es mayor.

Además, la barrera de energía libre varía de altura con las distintas mutaciones. Los cuatro mapas descritos anteriormente provocan en todos los casos un aumento de la barrera de energía libre con respecto a la que presenta 1PGB. Adicionalmente, como ya hemos explicado, el mínimo correspondiente al estado plegado se hace progresivamente más estrecho, es decir, tiene una menor variabilidad conformacional. Eliminando cuatro de los contactos presentes en 1PGB que no tienen un equivalente en 2GB1 (**mapa 1**), la barrera de energía libre se incrementa ligeramente. Activar los contactos de 2GB1 que no están en 1PGB (**mapa 2**) tiene el mismo efecto, aunque el aumento de la barrera es ligeramente superior que en el caso anterior. Cuando combinamos estas dos modificaciones en el **mapa 3**, la barrera también aumenta. Pero donde más se observa la diferencia entre unos mapas y otros es en el estrechamiento del mínimo que corresponde al estado plegado, como ya hemos explicado anteriormente. En el **mapa 4**, la barrera

no aumenta significativamente respecto a la que tiene lugar con el **mapa 3**, pero sí se nota una mejor definición del estado de baja energía.

A medida que disminuye el número de contactos entre las hebras y los lazos de 1PGB, la lámina deja de estar sobredefinida con respecto a las interacciones que estabilizan la hélice, y el estado de baja energía incluye una menor variedad de configuraciones. Haciendo un análisis de las conformaciones que dan lugar al estado de baja energía (o estado plegado) que aparece a T_m^* con los diferentes mapas, vemos que éstos presenta diferencias sustanciales de unos a otros. En la Figura 3.14 se representa la frecuencia con que aparece un contacto nativo entre todas las estructuras de baja energía resultantes del muestreo a T_m^* . En el caso de 2GB1, la estructura está prácticamente formada, dado que la mayoría de los contactos están formados con una frecuencia elevada. Sin embargo, el estado plegado que aparece a T_m^* cuando utilizamos el mapa de 1PGB es un estado en el que la hélice y los contactos de ésta con los otros elementos de estructura secundaria están formados con menor frecuencia. Es decir, este estado engloba configuraciones en las que la hélice está formada y configuraciones en las que no está bien definida. Es un estado de baja energía peor definido que el anterior. Al disminuir la temperatura, la hélice adquiere la estructura nativa. Este efecto ya lo hemos observado en el caso de utilizar el mapa resultante con un valor $d_{max} = 7.5 \text{ \AA}$ entre C_α . En este caso, sin embargo, no es lo suficientemente acusado como para dar lugar a un segundo pico, o un hombro, en la curva de capacidad calorífica. A medida que se va modificando el mapa de contactos de 1PGB, de manera que se reduce el número de contactos que estabilizan la lámina sobre la hélice, el mínimo que corresponde al estado plegado se estrecha (ver Figura 3.13), dando lugar a un estado de baja energía mejor definido donde los contactos de la hélice tienen una mayor frecuencia, como se aprecia en los resultados obtenidos usando los **mapas 1** a **4**. El tránsito se hace más brusco a medida que el estado plegado está mejor definido, es decir, la diferencia en estructura de un estado a otro es mayor y la barrera de energía libre aumenta. Cuando modificamos el mapa de contactos de manera que se eliminan

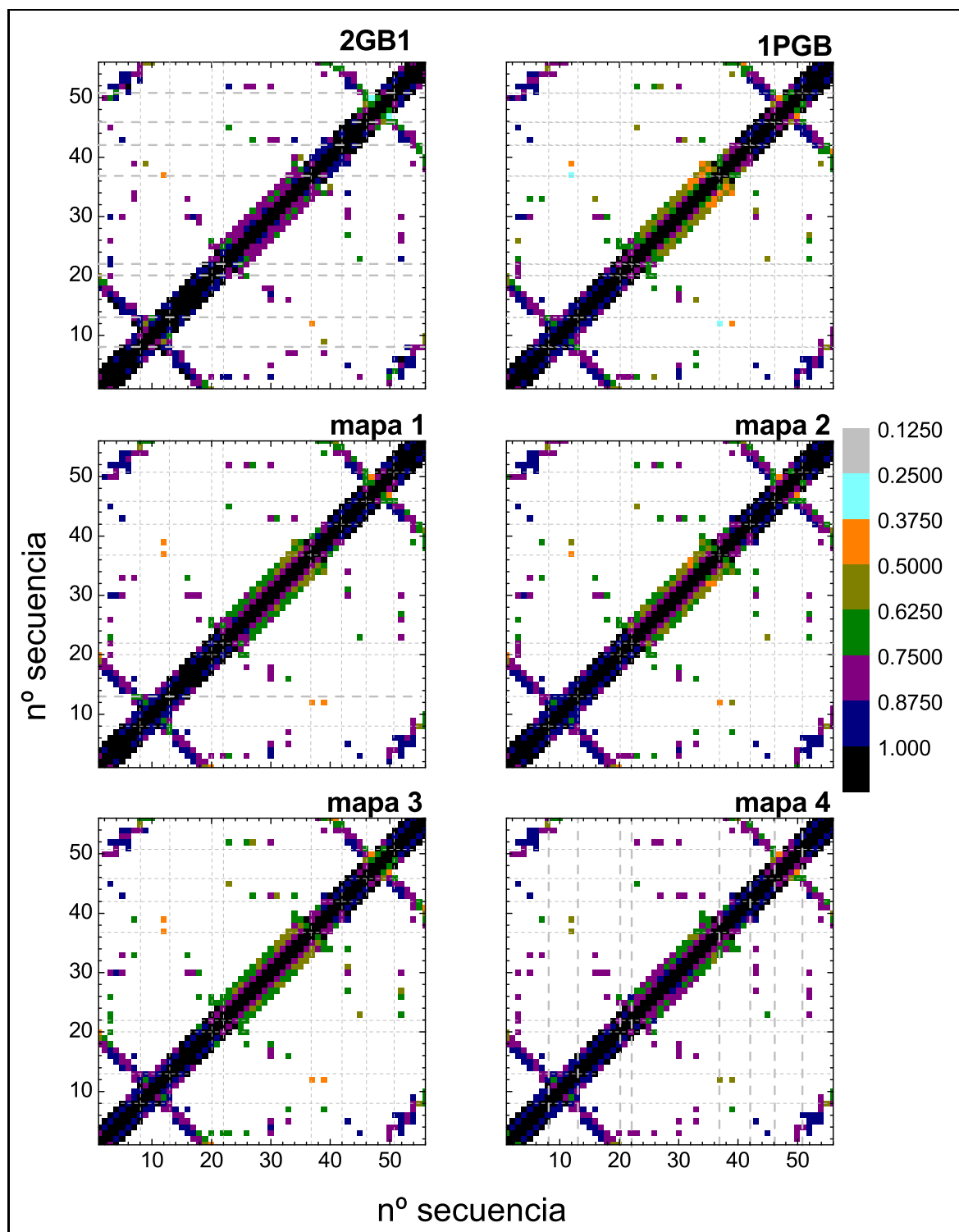


Figura 3.14: Frecuencias de los contactos nativos para el estado de baja energía que aparece a T_m^* . Un color más oscuro indica una mayor frecuencia.

los contactos que 1PGB tiene y 2GB1 no, y añadimos los contactos de interacción de la hélice con el resto de elementos de estructura secundaria, la transición se produce de manera más cooperativa. Es decir, tanto los contactos que el mapa 1PGB presenta de más con respecto al de 2GB1 como los que presenta de menos son responsables de que la transición tenga propiedades diferentes. De hecho, la barrera sigue siendo menor para todos los mapas que para el de 2GB1. Las pocas diferencias que todavía quedan entre el **mapa 4** y el de 2GB1 son suficientes para que las propiedades de la transición varíen. Por lo tanto, la distribución individual de todos los contactos tiene un efecto directo sobre la cooperatividad de la transición. Además, se puede deducir que una distribución más homogénea de los contactos entre todos los elementos de la estructura da lugar a una barrera de energía libre mayor.

No es el objetivo de este trabajo determinar cuál de los dos mapas experimentales es correcto, si es que puede decirse que alguno sea erróneo. Lo que queda claro de este estudio es lo determinante que es el mapa de contactos escogido para definir las interacciones del modelo. Queda claro en cualquier caso, a partir de los dos mapas (2GB1 y 1PGB), que la transición de plegamiento de esta proteína es de primer orden. Sin embargo, pequeñas diferencias en el mapa, aunque las características generales de la transición se mantienen, tienen influencia en los detalles de las propiedades termodinámicas del tránsito.

3.6. Resumen y conclusiones del Capítulo.

En este capítulo hemos estudiado cómo los distintos parámetros de un potencial basado en la topología del estado nativo influyen en la transición de plegamiento resultante de la simulación. Para ello, hemos utilizado un modelo simplificado en el que los aminoácidos se reducen a un único centro de interacción situado en la posición del C_α . Hemos definido un potencial muy sencillo que depende sólo de dos parámetros: la distancia de corte, o d_{max} , que define qué contactos estabilizan el estado nativo; y la

anchura del potencial, a , que determina el intervalo de acción de la atracción entre pares de aminoácidos.

Hemos comprobado la enorme importancia que el valor de d_{max} tiene en las propiedades de la transición de plegamiento simulada, dado que determina el número y situación de los contactos nativos en los que se basa el potencial. Hemos visto cómo un valor de d_{max} demasiado bajo crea, de manera artificial, zonas de menor estabilidad con respecto de otras. Esto provoca la aparición de intermedios termodinámicos que no se corresponden con los datos experimentales. Sin embargo, una vez que se elige un valor adecuado para la distancia de corte, los resultados que obtenemos demuestran que con modelos muy sencillos basados en un único centro de interacción podemos simular de manera correcta transiciones termodinámicas de primer orden que coinciden, al menos cualitativamente, con los datos experimentales. No es necesario, por tanto, utilizar representaciones más complejas de las proteínas para poder estudiar el proceso de plegamiento con potenciales de tipo $G\bar{o}$, como se había sugerido anteriormente^{57,73}.

Dada la importancia que tiene d_{max} a la hora de trabajar con este tipo de potenciales y que el valor más apropiado depende de la estructura concreta, hemos buscado la forma de determinar cuál es ese valor. Para ello, utilizamos como referencia un mapa de contactos basado en el empaquetamiento de todos los átomos pesados que forman parte de los aminoácidos. De esta manera, comparando los mapas que se obtienen con distintos valores de d_{max} con el mapa de contactos considerado como “real”, se puede determinar el valor de distancia de corte más adecuado a cada estructura. Por otra parte, d_{max} es una distancia de corte entre C_α . Sin embargo, podemos utilizar directamente el mapa de contactos basado en distancias entre átomos pesados para determinar los pares que forman contactos en el estado nativo. De esta forma, tenemos un criterio independiente de la estructura para establecer el mapa de contactos que define la topología del estado nativo, y por lo tanto el potencial de nuestro modelo.

No menos importante para las propiedades de la transición de plegamiento es el

valor de a , que representa la desviación respecto a las distancias nativas que se consideran dentro del alcance del potencial de atracción. En el intervalo de valores que hemos analizado, entre 0.3 y 1.0 Å, la variación en la barrera de energía libre entre el estado plegado y el desnaturalizado es grande, aún tratándose de un intervalo de valores bastante pequeño. Para el valor de a más elevado, dicha barrera incluso desaparece, dando lugar a una transición gradual. Se pasa de obtener transiciones de dos estados muy abruptas cuando se utilizan valores pequeños de a , a la desaparición de la barrera de energía libre con potenciales anchos. Queda patente entonces la relevancia de elegir adecuadamente el valor de a , sobre todo dadas las implicaciones que tiene en el camino y la dinámica del plegamiento. En estudios sobre la dinámica del plegamiento que se hacen utilizando potenciales basados en la topología del estado nativo, revisados en el Capítulo 1, las diferentes formas del potencial que se utilizan (normalmente de tipo Lennard-Jones con distintos exponentes que controlan la anchura) pueden dar lugar a un tipo de transición u otro. Sin embargo, este tipo de análisis termodinámico no se suele hacer. No quiere decir esto que los estudios realizados hasta el momento con potenciales de tipo Gō sean incorrectos, pero sería adecuado incluir este tipo de análisis para poder garantizar con mayor seguridad el carácter predictivo que puedan tener estos modelos.

Las variaciones que hemos observado en la barrera de energía libre en nuestras simulaciones se deben a la influencia que la anchura de potencial tiene en el estado plegado y, sobre todo, en el estado desnaturalizado. Dado que estamos estudiando un potencial basado en la topología del estado nativo, es sorprendente comprobar la gran variación observada en las propiedades del estado desnaturalizado. Con este tipo de potenciales, normalmente se ignora el estado desnaturalizado, ya que se considera que no puede estar bien definido. Sin embargo, hemos observado que su energía y, principalmente, su entropía varían al utilizar potenciales más o menos anchos, lo cual se traduce en una variación en las propiedades termodinámicas de la transición misma y afecta directamente a la estabilidad del modelo de proteína.

Los resultados obtenidos para las diferentes anchuras de potencial nos han permitido establecer el valor más apropiado de a para continuar con el estudio. Evidentemente, es necesario elegir un valor que responda a las observación experimentales. Por lo tanto, aquellos valores de a que no dan lugar, para la proteína que se está utilizando en este estudio, a una transición de dos estados, no son apropiados. Por otra parte, tampoco es adecuado utilizar potenciales demasiado estrechos, que dan lugar a transiciones muy abruptas. Combinando una anchura de potencial de $a = 0.7 \text{ \AA}$ con el mapa de contactos definido a partir de todos los átomos pesados de la proteína, observamos que los resultados son plenamente concordantes con los resultados experimentales existentes para la proteína empleada como referencia.

Una vez estudiados los parámetros del potencial, de manera que éste sea apropiado para el estudio de las propiedades de la transición de plegamiento, hemos visto que el mapa de contactos concreto utilizado tiene también una gran influencia en dichas propiedades. Esto es lógico dado que se trata de un potencial basado en la topología del estado nativo. Sin embargo, si bien es cierto que dos mapas con características generales de la estructura similares dan lugar al mismo tipo de transición, pequeñas variaciones en el mapa de contactos empleado influyen en la barrera de energía libre y en la definición de los estados que intervienen. De hecho, dos mapas de contactos con ciertas diferencias que corresponden a la estructura de la misma proteína, determinada por técnicas experimentales diferentes, dan lugar al mismo tipo de plegamiento: una transición de dos estados. Pero la barrera de energía libre de las dos transiciones es diferente. Hemos planteado la posibilidad de que una distribución homogénea de los contactos sea la responsable de una mayor cooperatividad en el tránsito. La pregunta que se plantea entonces es: ¿qué características de la estructura determinan el tipo de transición termodinámica que presentan las proteínas en su plegamiento? En el siguiente capítulo usamos el modelo descrito aquí para intentar dar respuesta a esta pregunta.

Capítulo 4

Topología y transición de plegamiento de proteínas.

Como comentamos en el Capítulo 1, la adquisición de estructura terciaria de muchas proteínas con un único dominio es un proceso termodinámico de dos estados que transcurre de manera cooperativa. Las desviaciones de este comportamiento se pueden interpretar de diferentes maneras. En el mismo capítulo hemos mencionado también que, en el caso del plegamiento ultra-rápido de algunas proteínas se ha establecido una relación entre la velocidad de plegamiento y la posibilidad de una barrera de energía libre muy pequeña o prácticamente inexistente¹³⁻¹⁷. Aunque la mayoría de las proteínas parecen haber sido seleccionadas por la evolución para presentar un plegamiento de dos estados, desde un punto de vista químico-físico es posible un tipo de plegamiento sin barreras (también llamado de tipo *downhill*)²¹⁻²⁵. En este capítulo tratamos este problema desde una perspectiva computacional utilizando el potencial basado en la topología del estado nativo cuyas características hemos estudiado en el capítulo anterior. Utilizar este tipo de potenciales para estudiar las propiedades termodinámicas del plegamiento es adecuado^{95,141} dada la relación que hay entre las barreras de energía y las velocidades de plegamiento, y el hecho de que estas últimas presentan una correlación importante

con el tamaño²⁴ y la complejidad estructural de las proteínas³⁸.

En este capítulo estudiamos qué características de la topología del estado nativo, si es que hay alguna en concreto, determinan las propiedades termodinámicas de las transiciones de plegamiento. Para ello vamos a utilizar el potencial de tipo Gō descrito en el Capítulo 2. En el capítulo anterior hemos visto cómo la definición del potencial mismo tiene una gran influencia en la presencia o no de una barrera de energía libre entre el estado nativo y el desnaturalizado y, en caso de existir esta barrera, en su altura^{75,129}. Además, en dicho capítulo hemos podido comprobar que el mapa de contactos tiene una influencia directa sobre las propiedades de la transición. La distribución de los contactos en el mapa es un factor crucial a la hora de describir la transición de plegamiento pero, además, es importante tener en cuenta la relación entre los diferentes tipos de interacciones que intervienen: locales y de largo alcance^{76,142}. El papel que estas interacciones juegan en el plegamiento de proteínas ha sido estudiado, sobre todo, desde un punto de vista cinético. En este capítulo estudiamos la influencia que tienen las distintas contribuciones a la energía de estos dos tipos de interacciones en la termodinámica del plegamiento. Para ello hemos variado la intensidad relativa de las interacciones de largo y corto alcance y hemos comprobado cuál es el efecto de esta modificación en las características del tránsito simulado. Podemos estudiar así qué importancia tienen los dos tipos de interacciones en el proceso de plegamiento. Si bien hay autores que afirman que son las interacciones de largo alcance las que son cruciales para el plegamiento de proteínas⁵⁴, otros autores opinan que la estructura nativa está determinada por las interacciones locales y que las no locales sólo proporcionan cierto grado de estabilidad⁴. Sin embargo, nuestros resultados muestran que es necesario cierto equilibrio entre ambos tipos de interacciones para que el plegamiento tenga un comportamiento cooperativo¹⁴². Aunque en trabajos anteriores ya se había observado este hecho^{75,76}, el análisis llevado a cabo en la primera parte de este capítulo nos permite profundizar en los motivos que provocan la aparición de una barrera de energía libre a lo largo del

proceso de plegamiento.

En una segunda parte del capítulo hemos llevado a cabo un estudio sobre el plegamiento de distintas proteínas con diferentes tamaños y topologías. Comparar los resultados obtenidos con los datos experimentales disponibles nos permite comprobar si las diferentes características de las topologías explican los distintos tipos de transiciones que tienen lugar. Además, podremos evaluar qué influencia tienen los dos tipos de interacciones considerados sin que la proporción relativa entre ambos se deba a modificaciones artificiales, sino a la estructura misma. Por último, hemos realizado un estudio del efecto que tienen contactos concretos en la termodinámica del plegamiento.

4.1. Interacciones de largo y corto alcance.

A partir del estudio realizado en el capítulo anterior, sabemos que el potencial de tipo Gō utilizado, con una anchura de potencial $a = 0.7 \text{ \AA}$ y el mapa de contactos resultante de considerar el empaquetamiento de todos los átomos de las cadenas laterales, da buenos resultados a la hora de describir la transición de plegamiento. La función de potencial utilizada es la de la ecuación (2.1). Como ya hemos dicho, el valor de w_{ij} controla la profundidad del potencial, es decir, la atracción entre los aminoácidos. Dicho parámetro es distinto de cero para todos los contactos de tipo 1-3 y 1-4, y para aquellos pares de residuos que forman un contacto terciario en el estado nativo, recogidos en el mapa de contactos. En el estudio del capítulo anterior el valor de $w_{ij} = 1/a^2$ para todos aquellos casos en que $w_{ij} \neq 0$. Pero para poder estudiar el papel que juegan los distintos tipos de interacciones, hemos variado el valor de w_{ij} de forma que controlamos la contribución de cada una de estas interacciones a la energía total.

En el Capítulo 2 se han explicado los distintos tipos de interacciones que tenemos en cuenta en nuestro modelo:

- Por un lado están las interacciones de corto alcance, responsables de la rigidez

intrínseca de la cadena polipeptídica. En este grupo incluimos tanto las interacciones que hemos llamado de ángulo de enlace virtual o interacciones 1-3 y las de torsión o interacciones 1-4. Para todas estas interacciones se considera un valor $w_{ij} = w^l$.

- Las interacciones de largo alcance incluyen al resto de interacciones que se dan entre aquellos residuos que forman un contacto nativo según el mapa de contactos definido. Para estos contactos $w_{ij} = w^{lr}$ y, dependiendo de su valor, la proteína tendrá mayor o menor tendencia a colapsar.

La energía total viene definida entonces por la ecuación:

$$E^* = w^l \left[\sum_{i=2}^{N-1} u_{i-1, i+1} (r_{i-1, i+1}) + \sum_{i=2}^{N-2} u_{i-1, i+2} (r_{i-1, i+2}) \right] + w^{lr} \sum_{m=1}^{N_{lr}} u_m (r_{ij}) \quad (4.1)$$

donde N es el número de aminoácidos de la proteína, N_{lr} es el número de contactos nativos de largo alcance y $u_{ij} = (r_{ij} - d_{ij}^{nat})^2 - a^2$, truncado como hemos comentado en el Capítulo 2. Cuando $w^l = w^{lr} = 1/a^2$, el mínimo de energía es el mismo para todas las interacciones ($u_{ij}(d_{ij}^{nat}) = -1$) y la energía del estado nativo ideal es $E_N^* = -(N_l + N_{lr}) = -N_{con}$ (siendo N_l el número total de contactos locales; N_{lr} el número de contactos de largo alcance y N_{con} el número total de contactos presentes en el estado nativo). Cambiando los valores de w^l y w^{lr} variamos la intensidad relativa de los dos tipos de interacciones. Para poder comparar los resultados obtenidos con distintas combinaciones de w^l y w^{lr} , hemos asignado valores a estos parámetros de manera que la energía total del estado nativo sea la misma en todos los casos. De esta forma podemos estudiar la importancia de las interacciones de largo y corto alcance en la termodinámica del plegamiento de proteínas sin que esto afecte a la definición del estado nativo en términos de energía total.

Desde luego, el criterio elegido para decidir qué interacciones son locales y cuáles de largo alcance es un tanto arbitraria. De hecho, en la bibliografía se pueden encontrar

otros criterios⁹⁵. Pero dado que existe una relación entre la intensidad de las interacciones locales y la rigidez de la cadena, hemos decidido no incluir aquellos contactos con $|j-i| > 3$ dentro del conjunto de interacciones locales, para así poder controlar la rigidez de manera independiente a la estructura de la proteína.

4.2. Influencia de la rigidez de cadena en la termodinámica del plegamiento.

Para realizar este análisis hemos utilizado la misma proteína que en el capítulo anterior, cuyo código PDB es 2GB1. Los distintos valores que hemos dado a w^l y w^{lr} aparecen en la Tabla 4.1. Es importante tener en cuenta que la propia topología del estado nativo impone una proporción determinada de interacciones de largo y corto alcance. En este caso, la proteína tiene 107 contactos locales (54 corresponden a contactos de tipo 1-3 y 53 a interacciones 1-4) y 108 contactos de largo alcance. En otras estructuras, el número de contactos y la proporción respecto del total será diferente, lo que deberá tenerse en cuenta para definir las distintas contribuciones.

Para cada combinación de los parámetros w^l y w^{lr} hemos realizado varias simulaciones con el potencial definido en la sección anterior de la manera ya descrita en el Capítulo 2. Para el caso en que $w^l = w^{lr}$, $w = 1/(0.7)^2 = 2.04 \text{ \AA}^{-2}$, cuyos resultados hemos estudiado en el capítulo anterior. A partir de éstos y los resultados obtenidos de las simulaciones realizadas para las distintas contribuciones a la energía, para poder visualizar el diferente comportamiento del sistema, lo primero que hemos calculado son las correspondientes curvas de capacidad calorífica. A partir de estas curvas, representadas en la Figura 4.1, podemos observar que el plegamiento se produce a temperaturas mayores cuando la contribución de las interacciones de largo alcance es mayor. Esto quiere decir que, a medida que las interacciones de largo alcance adquieren mayor importancia en la energía del estado nativo, la estabilidad de la proteína aumenta. Dado

Tabla 4.1: Valores de w^l y w^{lr} utilizados en este estudio y la correspondiente contribución de las interacciones de largo alcance a la energía total del estado nativo ($\%E_{lr}^*$) para la proteína 2GB1.

| $w^l/\text{\AA}^{-2}$ | $w^{lr}/\text{\AA}^{-2}$ | $\%E_{lr}^*$ |
|-----------------------|--------------------------|--------------|
| 3.57 | 0.52 | 13 |
| 3.06 | 1.03 | 25 |
| 2.55 | 1.54 | 38 |
| 2.04 | 2.04 | 50 |
| 1.53 | 2.55 | 63 |
| 1.02 | 3.05 | 75 |
| 0.51 | 3.56 | 88 |
| $\simeq 0.$ | 4.06 | 100 |

que la energía total del estado nativo permanece constante de unos casos a otros, la variación en la rigidez de la cadena, que disminuye al aumentar $\%E_{lr}^*$, debe afectar a otras propiedades como pueden ser la entropía del estado plegado y el desnaturalizado, o la energía de este último estado.

Pero además, la variación de la contribución relativa de las distintas interacciones tiene un efecto importante sobre la forma misma de la curva de capacidad calorífica. Lo primero que se observa a este respecto es que, cuando la energía del estado nativo tiene una contribución predominante de uno de los tipos de interacciones, el pico de C_v^* se ensancha. Este ensanchamiento va acompañado de la aparición de un hombro, y en algunos casos incluso aparece un segundo pico, lo que indica que se trata de una transición de tres estados a través de un intermedio termodinámico.

Para los casos en que $\%E_{lr}^*$ es muy pequeño, la elevada rigidez de cadena dificulta el colapso de la misma, lo que sólo ocurre a temperaturas extremadamente bajas. A mayores temperaturas se forma (o deshace) la estructura secundaria de manera gradual. Debido a ello tiene lugar la aparición de los intermedios indeseados (ya que no se

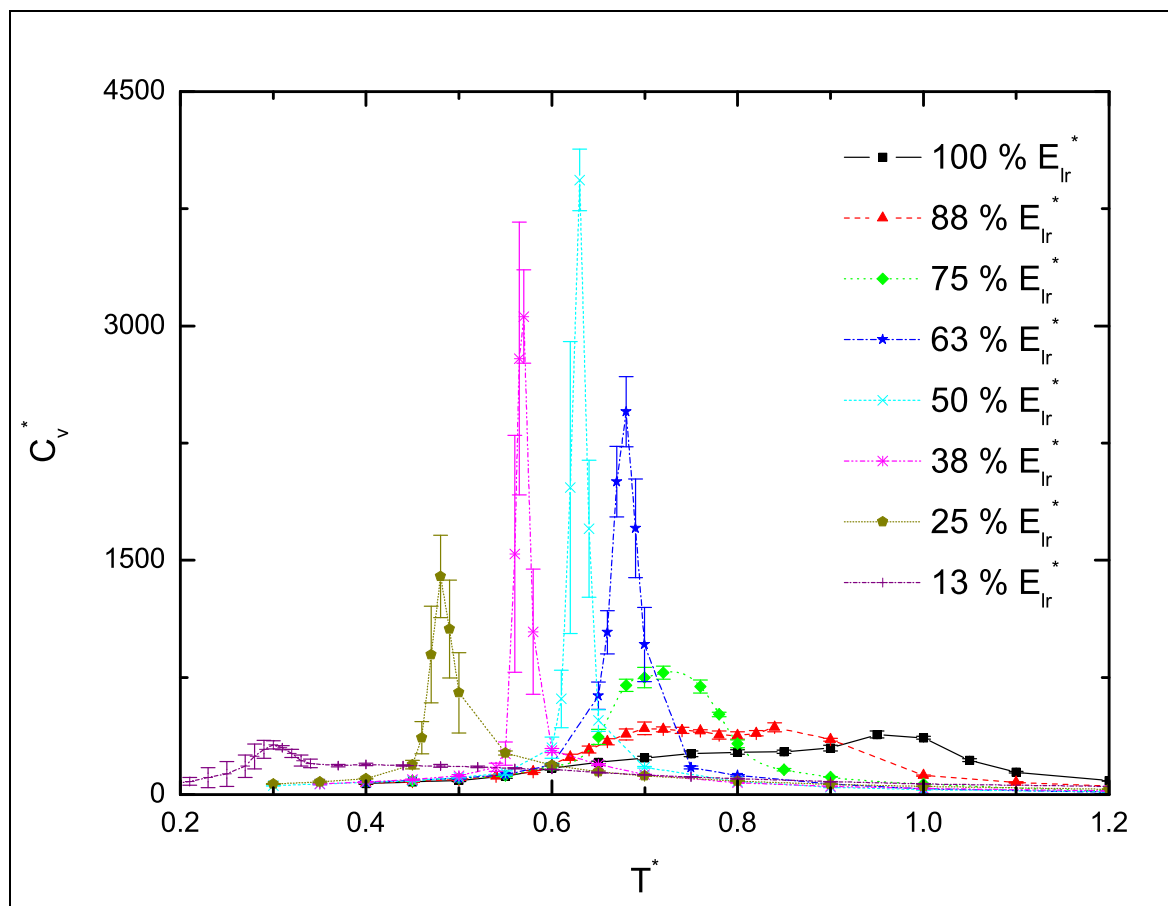


Figura 4.1: Curvas de capacidad calorífica obtenidas con las distintas contribuciones de las interacciones de largo alcance a la energía total del estado nativo.

observan experimentalmente para esta proteína). En los casos en que w^l tiene valores muy pequeños ($\%E_{lr}^*$ es grande) la cadena es muy flexible y el colapso de la misma se produce a temperaturas elevadas, lo cual dificulta la formación de la estructura secundaria a temperaturas menores. Esto también provoca la aparición de intermedios termodinámicos. Con el objetivo de caracterizar la estructura de los intermedios que aparecen, hemos calculado la frecuencia con que cada contacto nativo está formado en el conjunto de conformaciones que tienen lugar, como ya hemos hecho en casos anteriores. Este análisis lo hemos llevado a cabo a la temperatura intermedia entre el pico y el hombro ($T^* = 0.37$ para $w^{lr} = 0.52 \text{ \AA}^{-2}$ y $\%E_{lr}^* = 13$) o entre los dos picos ($T^* = 0.78$ para $w^{lr} = 3.56 \text{ \AA}^{-2}$ y $\%E_{lr}^* = 88$), los resultados obtenidos aparecen en la Figura 4.2.

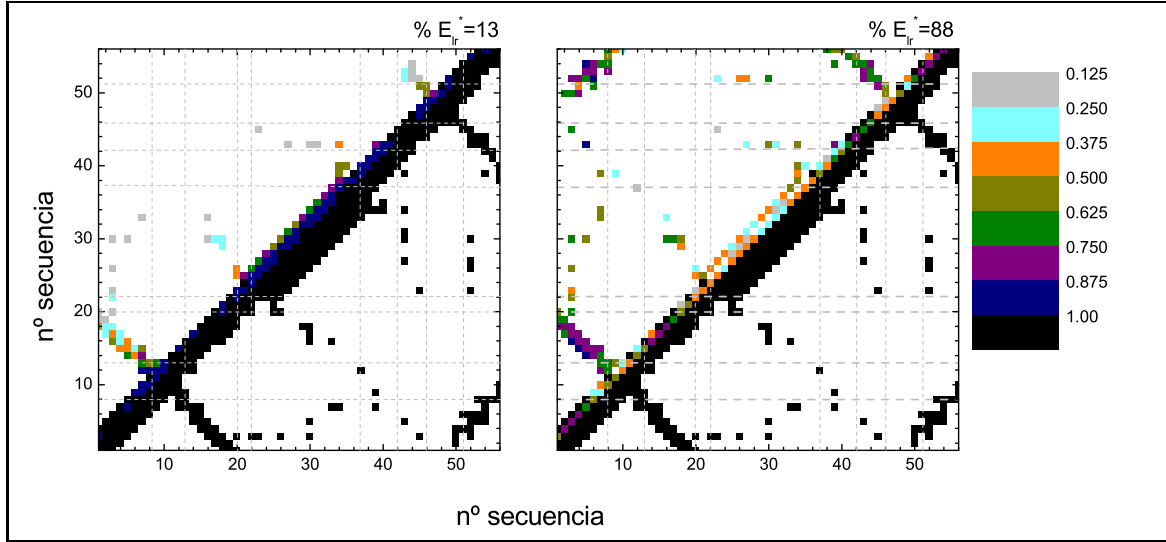


Figura 4.2: Mapa de frecuencia de los contactos nativos en las conformaciones que aparecen a las respectivas temperaturas con $w^{lr} = 0.52 \text{ \AA}^{-2}$ y $\%E_{lr}^* = 13$ ($T^* = 0.37$), a la izquierda, y con $w^{lr} = 3.56 \text{ \AA}^{-2}$ y $\%E_{lr}^* = 88$ ($T^* = 0.78$), a la derecha. La escala de colores indica la frecuencia de contacto en el triángulo superior izquierdo. En el triángulo inferior derecho aparecen los contactos del estado nativo.

Vemos que, efectivamente, la estructura media de estos intermedios corresponde con lo propuesto anteriormente:

- En el caso de w^{lr} bajo, a esa temperatura la estructura terciaria aún no se ha formado: los contactos de largo alcance tienen una baja frecuencia. A menor temperatura se produce el colapso y la proteína adquiere el estado nativo.
- Cuando w^{lr} tiene un valor elevado, el intermedio corresponde a una configuración colapsada: los contactos entre las hebras β tienen una frecuencia elevada. Sin embargo, los contactos locales se forman con poca frecuencia. Esto quiere decir que la hélice y los contactos de ésta con la lámina están por formar y sólo lo hacen al disminuir la temperatura.

Queda clara, por tanto, la relevancia de las distintas contribuciones a la energía del estado nativo por parte de las interacciones presentes en la termodinámica del plegamiento. Para continuar con el estudio, consideramos que el valor de T_m^* de cada caso estudiado

corresponde al valor de temperatura al que aparece el máximo absoluto de su capacidad calorífica.

Ya hemos visto en el capítulo anterior que para hacer un análisis más riguroso de las características termodinámicas de la transición hay que recurrir a los histogramas de energía y perfiles de energía libre. En la Figura 4.3 podemos ver dichos histogramas y perfiles a T_m^* para las distintas contribuciones a la energía del estado nativo que hemos analizado. De esta manera obtenemos una explicación al hecho de que el pico de capacidad calorífica se ensanche cuando $\%E_{lr}^*$ tiende a valores extremos: en estos casos, el histograma de energía deja de tener una distribución bimodal y pasa a ser unimodal, lo cual se traduce en una desaparición de la barrera de energía libre entre los estados plegado y desnaturalizado. La transición deja de ser de primer orden para convertirse en una transición sin barrera en la que se pasa de un estado a otro de manera gradual. Este tipo de transición desde luego no corresponde al plegamiento experimental de 2GB1, por lo que podemos decir que para obtener una descripción correcta de la transición que tiene lugar es necesario un equilibrio entre los distintos tipos de interacciones que intervienen.

Por otra parte, en la Figura 4.3 también se observa que la desaparición de la barrera de energía libre a medida que nos alejamos de las contribuciones intermedias se produce de manera diferente cuando se gana energía local en el estado nativo o si lo que priman son las interacciones de largo alcance. Así, cuando la contribución de la energía no local a la energía del estado nativo decrece del 50% al 13%, vemos que lo que sucede es que los dos mínimos se acercan, debido, principalmente, a una mayor estabilización del estado desnaturalizado (el pico correspondiente a este estado se desplaza hacia valores más negativos de energía). A pesar de que la barrera de energía libre permanece hasta valores muy bajos de w^{lr} , los dos mínimos van solapando más y más a medida que este parámetro disminuye, hasta que la barrera desaparece. Sin embargo, en el caso de aumentar la contribución de las interacciones de largo alcance, lo que ocurre es que

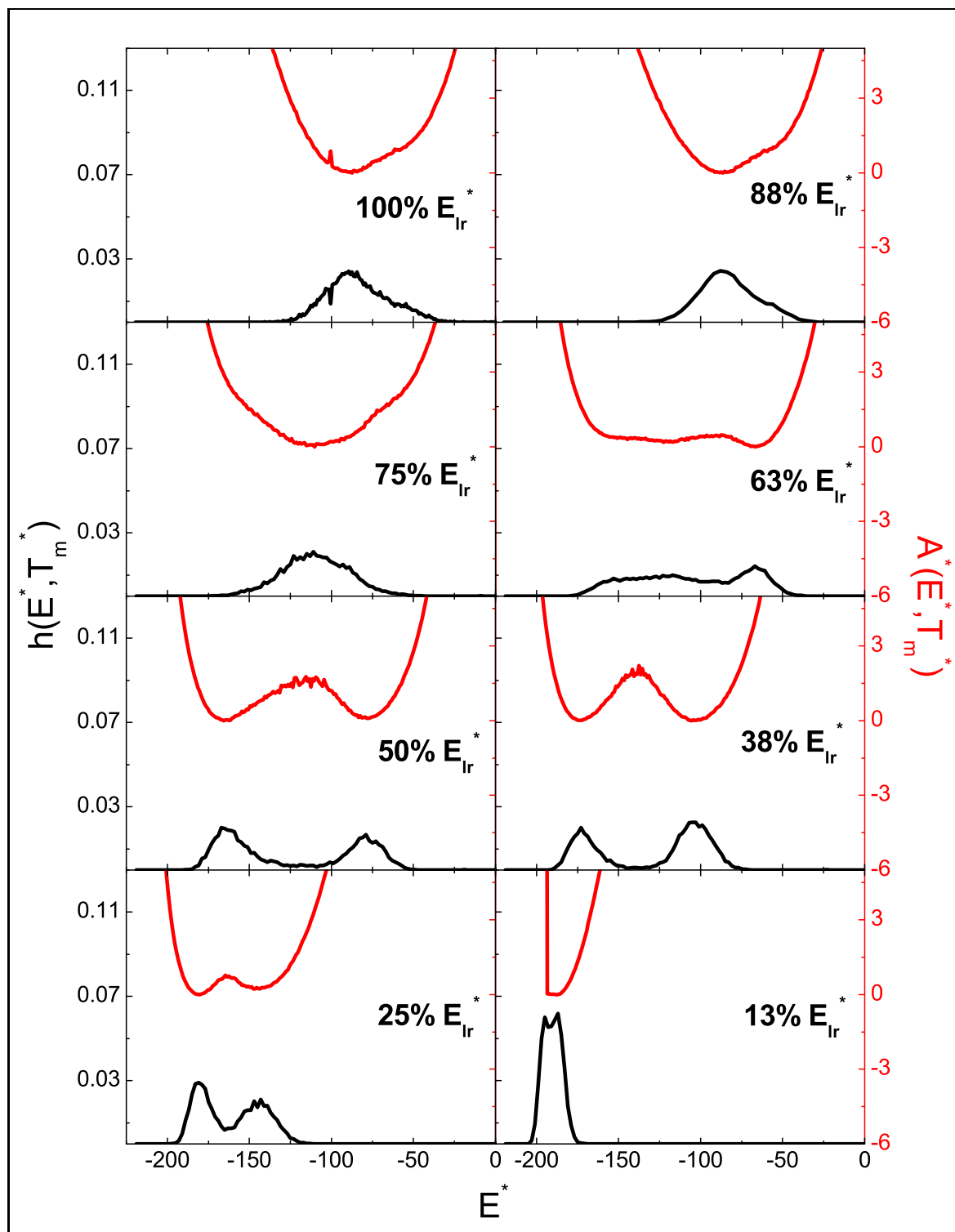


Figura 4.3: Histogramas de energía (negro) y perfiles de energía libre (rojo) para las distintas contribuciones de las interacciones de largo alcance a la energía del estado nativo ($\%E_{lr}^*$). La escala de los histogramas corresponde a los ejes izquierdos (en negro) y la escala de los perfiles de energía libre corresponde a los ejes derechos (en rojo).

la barrera disminuye progresivamente debido al ensanchamiento de los mínimos y, para valores de $\%E_{lr}^* \geq 75$, desaparece. Estos resultados dejan claro que tanto las interacciones locales como las de largo alcance son cruciales a la hora de dar lugar a una transición cooperativa.

Para poder explicar estas observaciones y profundizar en el proceso de plegamiento desde un punto de vista estructural hemos calculado los histogramas de contactos nativos formados a T_m^* de la Figura 4.4. En esta figura hemos separado los contactos locales de los de largo alcance. De esta manera podemos hacernos una idea del grado de estructura que se ha adquirido a esa temperatura en cada caso. Como ya hemos dicho anteriormente, para 2GB1 el número de contactos locales es 107 y el de contactos de largo alcance, 108. Cuando el pico del histograma aparece en torno a 100, podemos decir que la gran mayoría de los contactos nativos están formados a esa temperatura. Los histogramas de contactos, tanto locales como de largo alcance, para cadenas flexibles ($\%E_{lr}^* \geq 75$) presentan un único pico, al igual que sucede con los histogramas de energía total correspondientes de la Figura 4.3. Además, vemos que el número de contactos locales es muy bajo y que ni siquiera un número elevado de contactos de largo alcance induce la formación de estructura secundaria (es el caso de $\%E_{lr}^* = 75$). Para cadenas más rígidas que $\%E_{lr}^* = 75$, la distribución de contactos de largo alcance es bimodal, aunque las barreras presentes varían de manera importante de unos valores de $\%E_{lr}^*$ a otros. Sin embargo, la distribución de contactos locales presenta dos picos sólo para valores intermedios de rigidez, siendo de nuevo unimodal para cadenas muy rígidas. En estos últimos casos, prácticamente todos los contactos locales están formados a T_m^* , mientras que la estructura terciaria muestra un equilibrio entre configuraciones abiertas, con pocos contactos de largo alcance, y configuraciones compactas con un elevado número de contactos no locales formados. Este equilibrio se debe, principalmente, a la rotación de un único ángulo de torsión que corresponde a la zona de los giros de unión entre elementos de estructura secundaria. Los resultados evidencian que, en estos

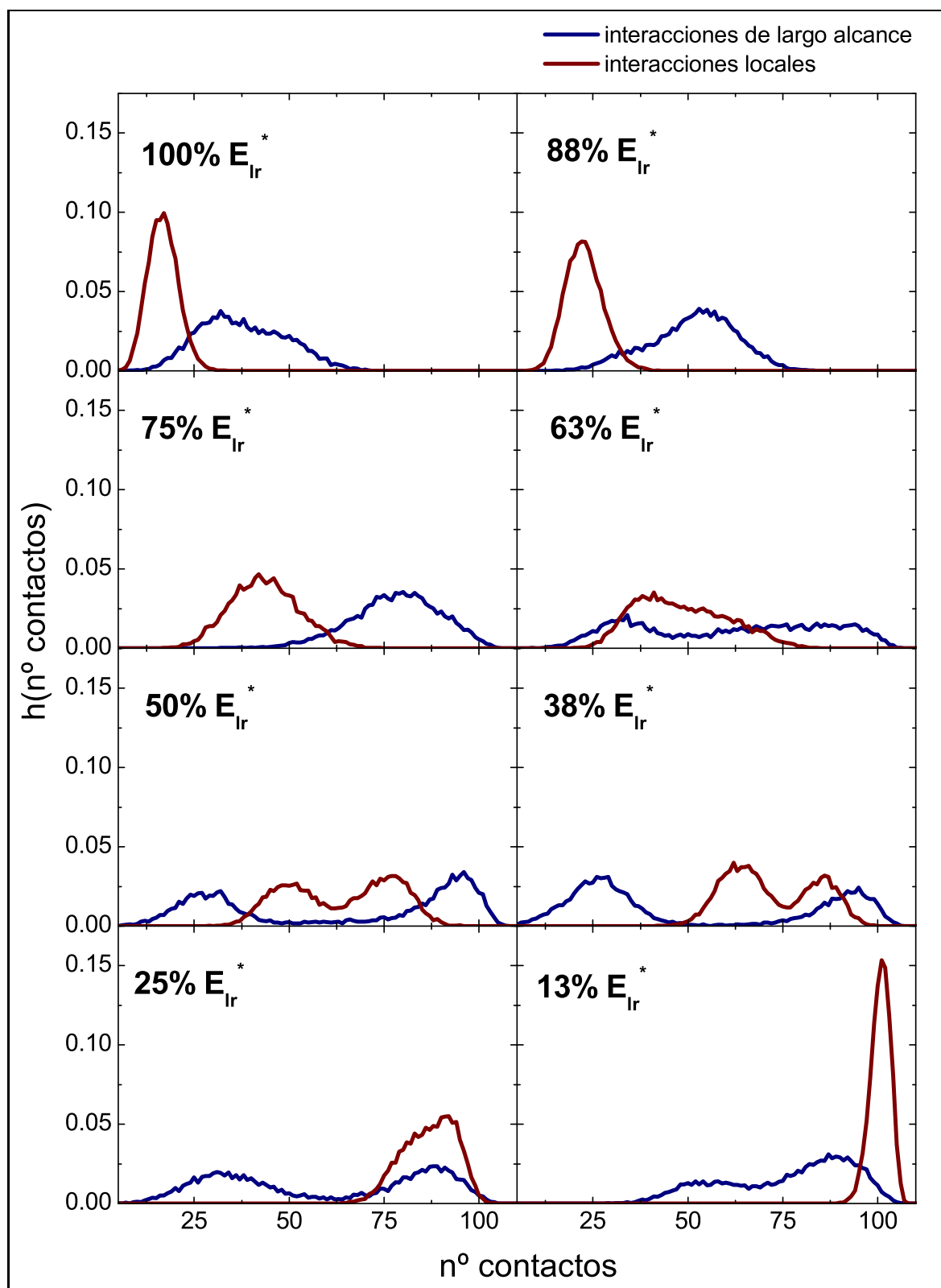


Figura 4.4: Histogramas de los contactos nativos de largo y corto alcance a T_m^* para las distintas contribuciones de las interacciones terciarias a la energía total del estado nativo.

casos de elevada rigidez, las interacciones locales dirigen el plegamiento, dando lugar a transiciones que, como ya hemos dicho anteriormente, no corresponden a las observaciones experimentales para el plegamiento de esta proteína. Tampoco se corresponden a estas observaciones experimentales los resultados obtenidos para cadenas muy flexibles (valores altos de $\%E_{lr}^*$). En estos casos, la distribución unimodal de los contactos locales hace pensar que éstos se forman de manera gradual. Para comprobar la veracidad de esta hipótesis en la Figura 4.5 hemos representado esta distribución a diferentes temperaturas. Vemos que, efectivamente, la distribución de contactos locales presenta un único pico que cambia de posición de unas temperaturas a otras. En esta misma figura podemos ver que el colapso y posterior reajuste de la estructura terciaria también se produce de manera gradual. Dado que ambas contribuciones presentan una distribución unimodal, la distribución de energía total presenta también un solo pico. En esta figura podemos también comprobar cómo la estructura secundaria está poco formada incluso a una temperatura inferior a T_m^* , mientras que a esta temperatura la estructura terciaria está formada en gran medida. Esto muestra que, para tener una transición de dos estados, las distintas interacciones que intervienen tienen que “cooperar” las unas con las otras de forma equitativa. En una transición cooperativa la formación de estructura secundaria y terciaria debe ser prácticamente simultánea, lo que no es posible cuando un tipo de interacción predomina sobre otro.

Hasta ahora hemos explicado cómo varía la transición termodinámica que tiene lugar con la rigidez de cadena. Pero hemos visto también en la Figura 4.1 que, además de las variaciones de la forma del pico de capacidad calorífica, éste se desplaza a temperaturas mayores a medida que aumenta la contribución de las interacciones de largo alcance. El estado nativo se vuelve más estable, a pesar de que su energía total es la misma en todos los casos. Por lo tanto, es necesario estudiar cómo varía la diferencia de energía y de entropía entre el estado plegado y el desnaturalizado de unos casos a otros para poder explicar estos cambios en la estabilidad de la proteína modelizada. De

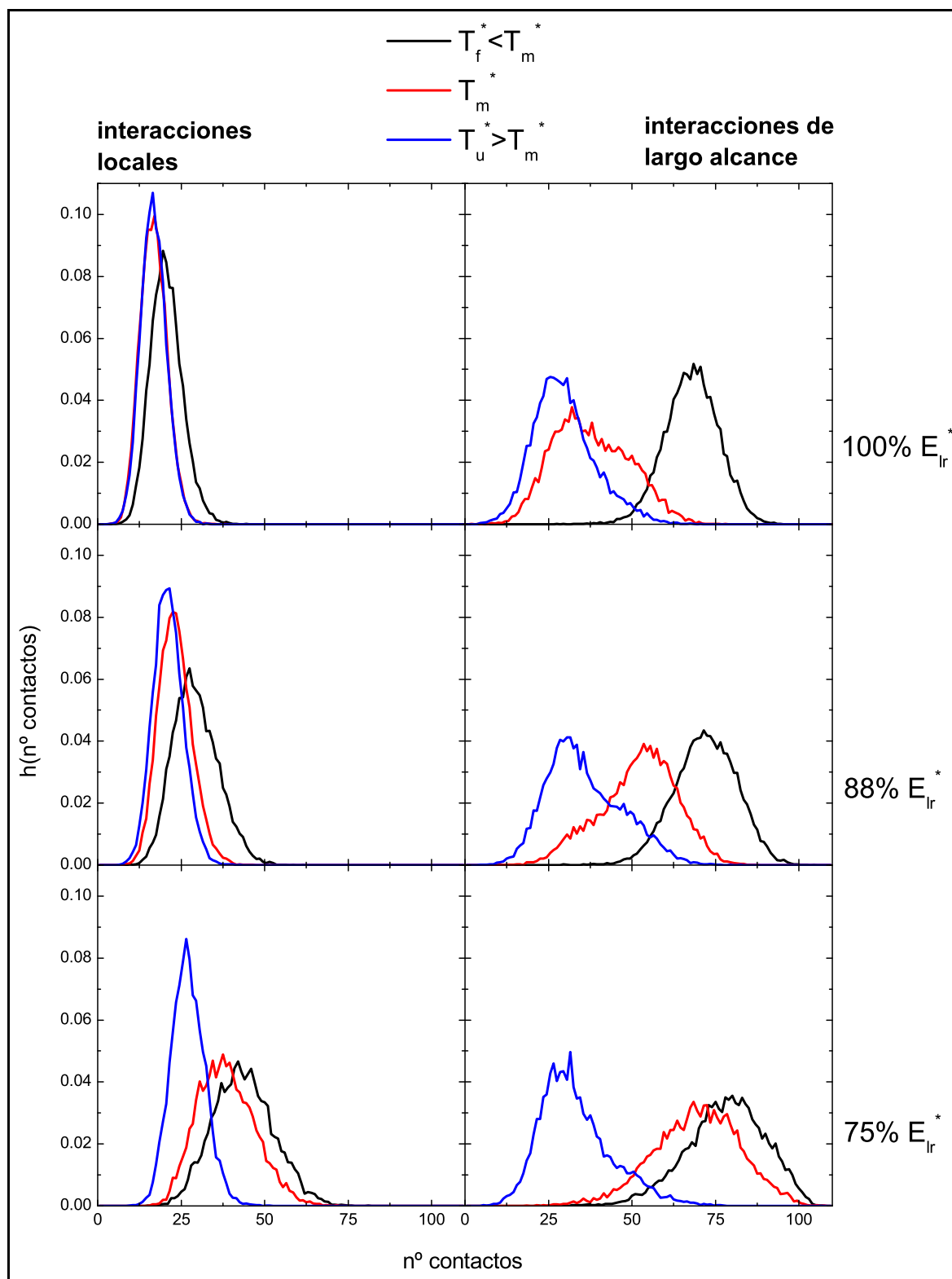


Figura 4.5: Histogramas de los contactos nativos de largo y corto alcance a las temperaturas indicadas para las cadenas flexibles.

Tabla 4.2: Variación de energía (ΔE_m^*) y entropía (ΔS_m^*) a la temperatura de transición para las distintas contribuciones de las interacciones de largo alcance ($\%E_{lr}^*$) que dan lugar a transiciones de dos estados.

| $\%E_{lr}^*$ | ΔE_m^* | ΔS_m^* | $T_m^* = \frac{\Delta E_m^*}{\Delta S_m^*}$ |
|--------------|----------------|----------------|---|
| 25 | 34 | 70 | 0.48 |
| 38 | 65 | 116 | 0.56 |
| 50 | 75 | 119 | 0.63 |
| 63 | 57 | 87 | 0.66 |

los resultados de nuestras simulaciones obtenemos directamente la distribución de energía y aplicando el método WHAM calculamos la distribución de entropía. Utilizando la ecuación (2.15), por integración de los diferentes picos correspondientes a cada estado, podemos calcular las diferencias mencionadas para los casos en que la transición es de dos estados, tal y como hemos explicado en el Capítulo 2. Los resultados obtenidos se muestran en la Tabla 4.2. Representando, además, estos valores frente a la contribución de energía no local, podemos observar que las tendencias de ΔE_m^* y de ΔS_m^* son similares (Figura 4.6). Al principio, a medida que $\%E_{lr}^*$ se incrementa, los valores de ΔE_m^* y de ΔS_m^* también aumentan. Esto sucede hasta que estas propiedades alcanzan un valor máximo. A partir de este punto, comienzan a disminuir. Sin embargo, en la Tabla 4.2 y en la Figura 4.1 vemos que el valor de T_m^* aumenta a lo largo de todo el intervalo de contribuciones. Dado que $T_m^* = \Delta E_m^* / \Delta S_m^*$, el incremento en ΔE_m^* es más importante que el de ΔS_m^* para contribuciones por debajo del máximo de las curvas de la Figura 4.6. Para valores mayores de $\%E_{lr}^*$ en que ΔE_m^* y de ΔS_m^* disminuyen, esta disminución es más notable para ΔS_m^* , lo que hace que también continúe aumentando la temperatura de transición.

Como ya hemos dicho, la variación en los en los resultados de unas contribuciones a otras se debe, principalmente, al efecto que tienen sobre la energía del estado desna-

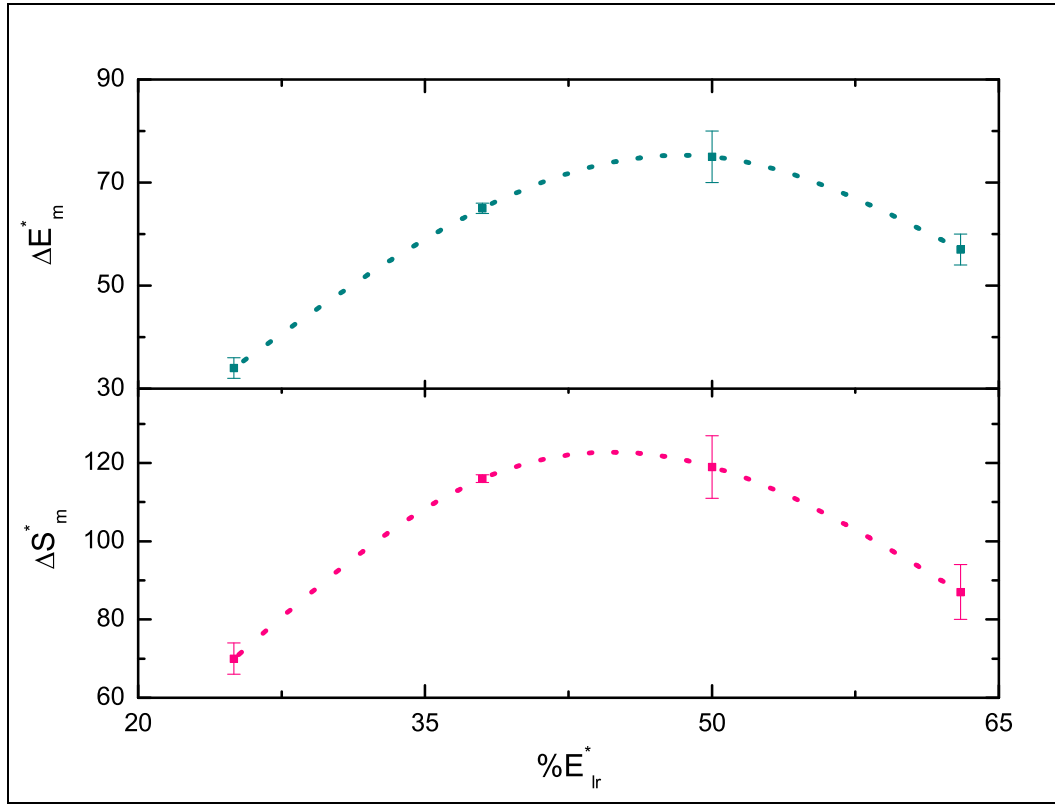


Figura 4.6: Valores de ΔE_m^* y ΔS_m^* a la temperatura de transición para los diferentes porcentajes de la energía de largo alcance respecto del total.

turalizado y la entropía de los estados plegado y desnaturalizado. Para los casos en que la cadena es más rígida, hemos visto que, a T_m^* , la cadena tiene un elevado grado de estructura secundaria (ver Figura 4.4). Esto hace que el estado desnaturalizado a T_m^* esté formado por un número de configuraciones más limitado que en casos de mayor flexibilidad. El estado desnaturalizado tiene pues una menor entropía que aquél al que da lugar una cadena más flexible. Esta disminución en la entropía del estado desnaturalizado provoca que $\Delta S_m^* = S_D - S_P$ (siendo S_D y S_P la entropía de los estados desnaturalizado y plegado, respectivamente) sea también menor. En el caso de que la contribución de las interacciones de largo alcance a la energía total sea elevada, la flexibilidad de la cadena confiere una mayor entropía al estado desnaturalizado. Pero además, esta flexibilidad se retiene también parcialmente en el estado de baja energía definido a T_m^* . En la Figura 4.3 vemos que el mínimo de energía libre correspondiente al estado plegado sufre un ensan-

chamiento considerable a medida que se pasa de cadenas más rígidas ($\%E_{lr}^* = 25$) a cadenas más flexibles ($\%E_{lr}^* = 63$). Este ensanchamiento del mínimo de energía libre se debe a que el estado plegado se vuelve más flexible y, por lo tanto, su entropía aumenta. Esto produce que la diferencia de entropía entre los estados plegado y desnaturalizado disminuya, a partir de cierto valor de $\%E_{lr}^*$, al aumentar la flexibilidad de la cadena. Cuando el potencial está en uno de los dos extremos del intervalo de contribuciones, ΔS_m^* disminuye considerablemente. Para que haya una barrera de energía entre ambos estados, la cadena debe tener una rigidez intermedia, es decir, las interacciones locales y de largo alcance deben estar compensadas.

Aunque el estado nativo tiene la misma energía en todos los casos a bajas temperaturas, la energía del estado plegado que aparece a T_m^* puede variar. Como muestra la Figura 4.6, la tendencia de ΔE_m^* es la misma que la de ΔS_m^* y la explicación la podemos encontrar en la Figura 4.3. Para cadenas flexibles, el pico correspondiente al estado de baja energía se va ensanchando a medida que aumenta $\%E_{lr}^*$. Esto se debe a un aumento de la entropía de ambos estados que ya hemos explicado. El ensanchamiento es tal que, para un valor de $\%E_{lr}^* = 63$, los picos solapan en gran medida de manera que disminuye ΔE_m^* . Para $\%E_{lr}^* \geq 75$ hay un único pico y $\Delta E_m^* = 0$. Esto significa que las interacciones de largo alcance provocan el colapso de la cadena con mayor facilidad a medida que su contribución a la energía total del estado nativo es mayor. Sin embargo, el colapso no siempre conduce directamente a la estructura nativa, sobre todo en el caso de cadenas muy flexibles en las que la estructura secundaria apenas se ha adquirido a T_m^* (ver Figura 4.4). En el caso de cadenas rígidas, vemos en la Figura 4.3 que la energía del estado desnaturalizado disminuye a medida que las interacciones locales adquieren una importancia mayor. Esto hace que ΔE_m^* disminuya al aumentar la rigidez de la cadena hasta que prácticamente se anula ($\%E_{lr}^* = 13$). La rigidez de la cadena hace que el estado desnaturalizado retenga un alto grado de estructura secundaria hasta temperaturas mayores que T_m^* , provocando que sea un estado con mucha energía residual. Al

tener el estado desnaturalizado menor energía (un valor más negativo) y mantenerse aproximadamente constante la del estado plegado, la diferencia de energía entre los dos estados disminuye cuanto mayor es la contribución de las interacciones locales.

Con este análisis podemos explicar por qué, para poder describir la transición del plegamiento de 2GB1 como una transición de primer orden, es necesario que haya un equilibrio entre las interacciones locales y las de largo alcance. Este equilibrio se traduce en una compensación entre la entropía y la energía de los estados que intervienen, que es lo que provoca la aparición de una barrera de energía libre.

Nuestros resultados muestran que, para 2GB1, hay un intervalo de contribuciones a la energía del estado nativo por parte de las diferentes interacciones en el que el plegamiento se produce a través de una transición de dos estados. Dicho intervalo comprende aproximadamente entre un 25 y un 63 % de energía de largo alcance en el estado nativo. En el caso particular de 2GB1, la proporción natural de contactos locales y de largo alcance es $\simeq 50\%$. Por lo tanto, cuando se toma el mismo valor de w_{ij} para ambos tipos de interacciones, nos encontramos en el intervalo en que la transición está correctamente definida. Para saber si ésta es una característica particular de la proteína que hemos elegido o, por el contrario, es algo general en las proteínas, hemos calculado cuál es la proporción de contactos de largo alcance respecto al total de contactos nativos para una base de datos de 1590 proteínas monoméricas no homólogas entre sí¹⁴³. A partir de los resultados obtenidos, en la Figura 4.7 representamos tanto el número de contactos de largo alcance como la proporción de estos contactos frente a la longitud de cadena. En la Figura 4.7a vemos que el número de contactos de largo alcance presenta una relación lineal con el número de aminoácidos de las proteínas. La pendiente de esta línea es 2.80 y el cuadrado del coeficiente de regresión lineal $r^2 = 0.98$. En la parte **b** de dicha figura, lo que se representa es la proporción de contactos de largo alcance con respecto al número total de contactos ($\%E_{lr}^*$). La mayoría de las proteínas monoméricas tienen un valor de $\%E_{lr}^*$ entre un 50 y un 60 %. En este intervalo, precisamente, el plegamiento de

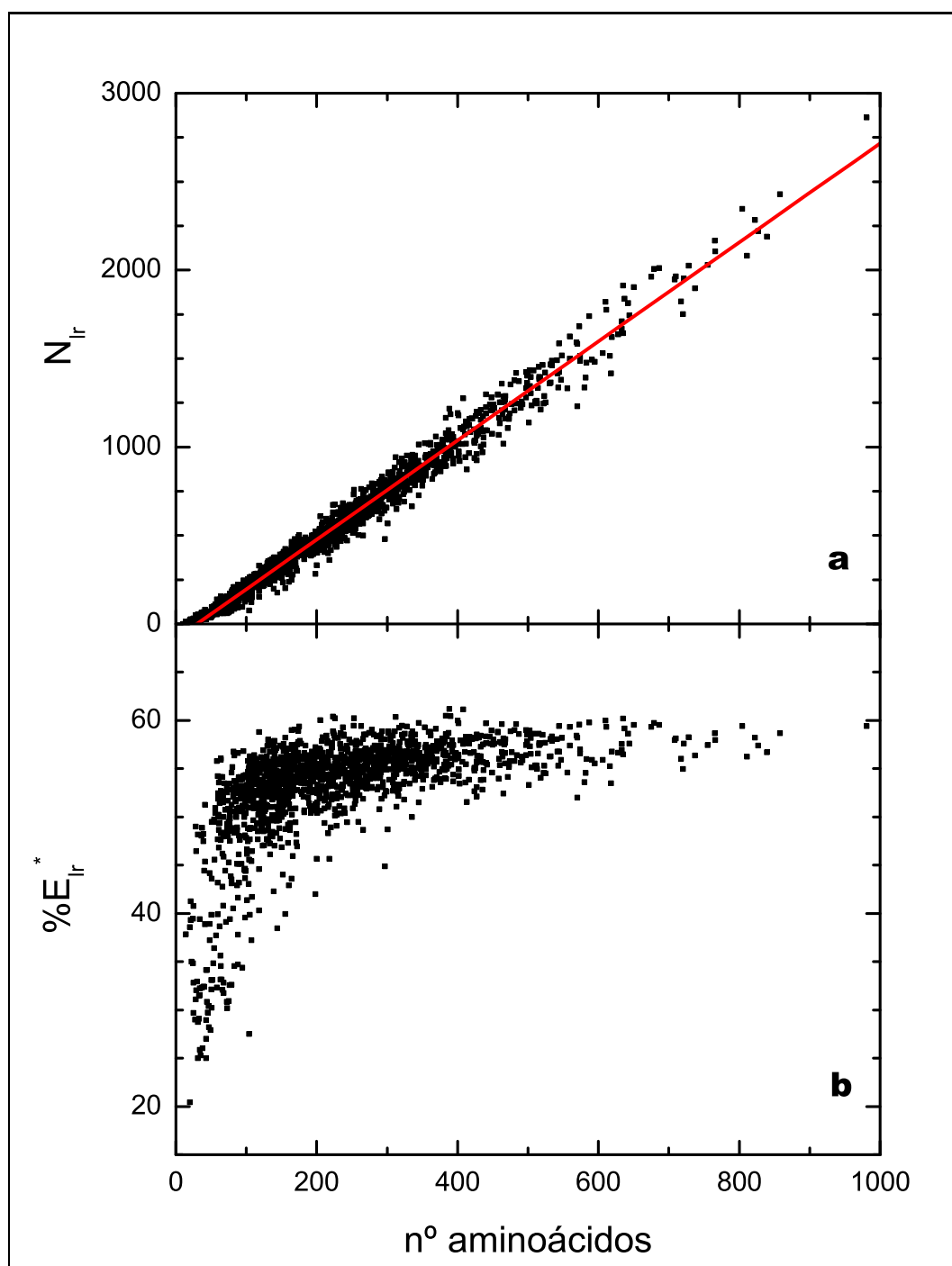


Figura 4.7: **a)** Número de contactos de largo alcance (N_{lr}) y **b)** proporción de estos contactos con respecto al total de contactos nativos, en función del número de aminoácidos para un conjunto de 1590 proteínas.

2GB1 se produce según una transición de dos estados. Estos resultados hacen pensar que la mayoría de las proteínas tienen, desde este punto de vista, un empaquetamiento similar y que, por lo tanto, presentan un equilibrio intrínseco entre las interacciones de largo alcance y las locales. Esta observación está de acuerdo con el hecho de que la cooperatividad parece ser una característica general del plegamiento de proteínas¹¹.

Sin embargo, en la Figura 4.7b podemos ver que, para algunas proteínas pequeñas, la proporción de contactos de largo alcance es sensiblemente menor que el 50 %. Si, como hemos visto, la distinta contribución a la energía del estado nativo da lugar a distintos tipos de transiciones, ¿presentan estas proteínas un comportamiento diferente?

Las proteínas que en la Figura 4.7 tienen una menor proporción de estructura terciaria presentan un porcentaje de interacciones de largo alcance por encima del 25 %. Aunque, según los resultados obtenidos para 2GB1, esto da lugar a una transición de dos estados, es probable que este intervalo no sea generalizable a todas las proteínas y que alejarse del 50 % de contactos terciarios sea una característica indicativa de otro tipo de transición termodinámica diferente a una de dos estados. Esto nos lleva a estudiar el comportamiento de un conjunto de proteínas cuyas características estructurales correspondan tanto a una proporción entre el 50 y el 60 % de contactos de largo alcance, como a un porcentaje inferior.

4.3. Transición termodinámica de proteínas con distinta topología.

Para este estudio hemos buscado proteínas pequeñas cuyas transiciones termodinámicas hayan sido caracterizadas experimentalmente. Además de incluir el dominio B1 de la proteína G de *Streptococcus* (**2GB1**) en el conjunto de proteínas estudiado, hemos elegido, en primer lugar, el complejo 2-oxoglutarato deshidrogenasa de *Escherichia coli*, cuyo código PDB es **1BBL**. Aunque existen algunos argumentos en contra²⁸⁻³³, el

plegamiento de esta proteína ha sido caracterizado como una transición sin barrera¹⁴.

En los trabajos que intentan demostrar que esta afirmación es errónea, se han utilizado diferentes mutantes y homólogos estructurales de 1BBL. Entre estos se encuentra el mutante E3BD F166W, cuyo código PDB es **1W4E** y cuya transición se considera un proceso de dos estados²⁸⁻³⁰. Por este motivo la hemos incluido en el conjunto de proteínas a estudiar. Esta proteína forma parte de los estudios de Naganathan y colaboradores sobre barreras de energía libre²⁵, donde también se ha caracterizado como una proteína con un plegamiento de dos estados. En este mismo trabajo se estudia también la transición del dominio periférico de la dihidrolipoamidil-acetiltransferasa del complejo piruvatodeshidrogenasa de *Bacillus stearothermophilus* (con código PDB **2PDD**) y la proteína de choque térmico de *Bacillus subtilis*, cuyo código PDB es **1CSP**. La primera de ellas, 2PDD, es un homólogo estructural de 1BBL y la forma silvestre de 1W4E. Se considera que su transición es un proceso de dos estados, pero con una barrera pequeña. La transición de 1CSP también se considera de tipo todo o nada.

Además, hemos incluido el inhibidor 2 de la quimotripsina, cuyo tránsito ha sido perfectamente caracterizado como un plegamiento con barrera y que se considera un paradigma del plegamiento de dos estados¹⁴⁴ (código **2CI2**).

Otra proteína que hemos tomado en consideración es el subdominio inicial de la villina, que tiene el código **1VII** en el PDB. Los datos experimentales apuntan a una transición de dos estados¹⁴⁵. Sin embargo, la elevada velocidad de plegamiento sugiere que, aunque existe una barrera de energía libre, ésta es muy pequeña. Los estudios teóricos hechos sobre esta proteína llegan a la conclusión de que se trata de un plegamiento muy poco cooperativo^{95,146}.

Por último, también hemos estudiado la miniproteína de la caja de triptófano (*Trp-cage miniprotein*), con sólo 20 aminoácidos y cuyo código PDB es **1L2Y**. Experimentalmente se piensa que su transición es de dos estados¹⁴⁷ pero los estudios teóricos apuntan a un plegamiento sin barrera⁹⁵.

La estructuras de estas proteínas, a las cuales nos referimos por su código PDB, aparecen representadas, con el correspondiente mapa de contactos, en la Figura 4.8. Podemos ver que estructuras homólogas, como son 1BBL y 2PDD, presentan mapas de contactos similares, aunque 2PDD tiene más contactos, entre otras cosas porque tiene una mayor longitud. Por otra parte, los mapas de 2PDD y 1W4E también se parecen mucho entre sí, lo cual es razonable teniendo en cuenta que 1W4E es un mutante de 2PDD. Dado que éstos son los mapas de contactos considerados en el potencial de interacción que estamos usando, es de esperar que las propiedades termodinámicas de la transición a que dan lugar sean también similares.

Como para 2GB1 el valor $w^l = w^{lr} = 1/a^2$ da resultados comparables con los experimentales, lo hemos utilizado para estudiar las transiciones termodinámicas de las proteínas que acabamos de presentar y analizar cómo su topología determina las características de las mismas.

En primer lugar, hemos estudiado la transición de 1BBL. Ésta es una proteína pequeña, 37 aminoácidos, que presenta una proporción relativamente baja de contactos de largo alcance. En total tiene 108 contactos, de los cuales 39 son contactos terciarios y 69 son locales (35 de tipo 1-3 y 34 de tipo 1-4). Esto implica que, para 1BBL, sólo un 36 % de la energía del estado nativo se debe a interacciones de largo alcance. Según los resultados obtenidos para 2GB1, este valor entra, aunque por poco, dentro del intervalo en el cual la transición resulta ser de primer orden. Por otra parte, en la Figura 4.7b vemos que un 36 % es un valor bastante alejado del 50 %, por encima del cual se encuentra la mayoría de las proteínas monoméricas de la base de datos estudiada. Para estudiar la transición de 1BBL y comprobar si, efectivamente, sus diferencias estructurales respecto a la tendencia más general provocan un comportamiento diferente, hemos llevado a cabo varias simulaciones con el potencial basado en la topología del estado nativo que estamos estudiando. Con el objetivo de caracterizar la termodinámica de su transición de plegamiento, hemos representado la curva de capacidad calorífica (Figura 4.9a) y la

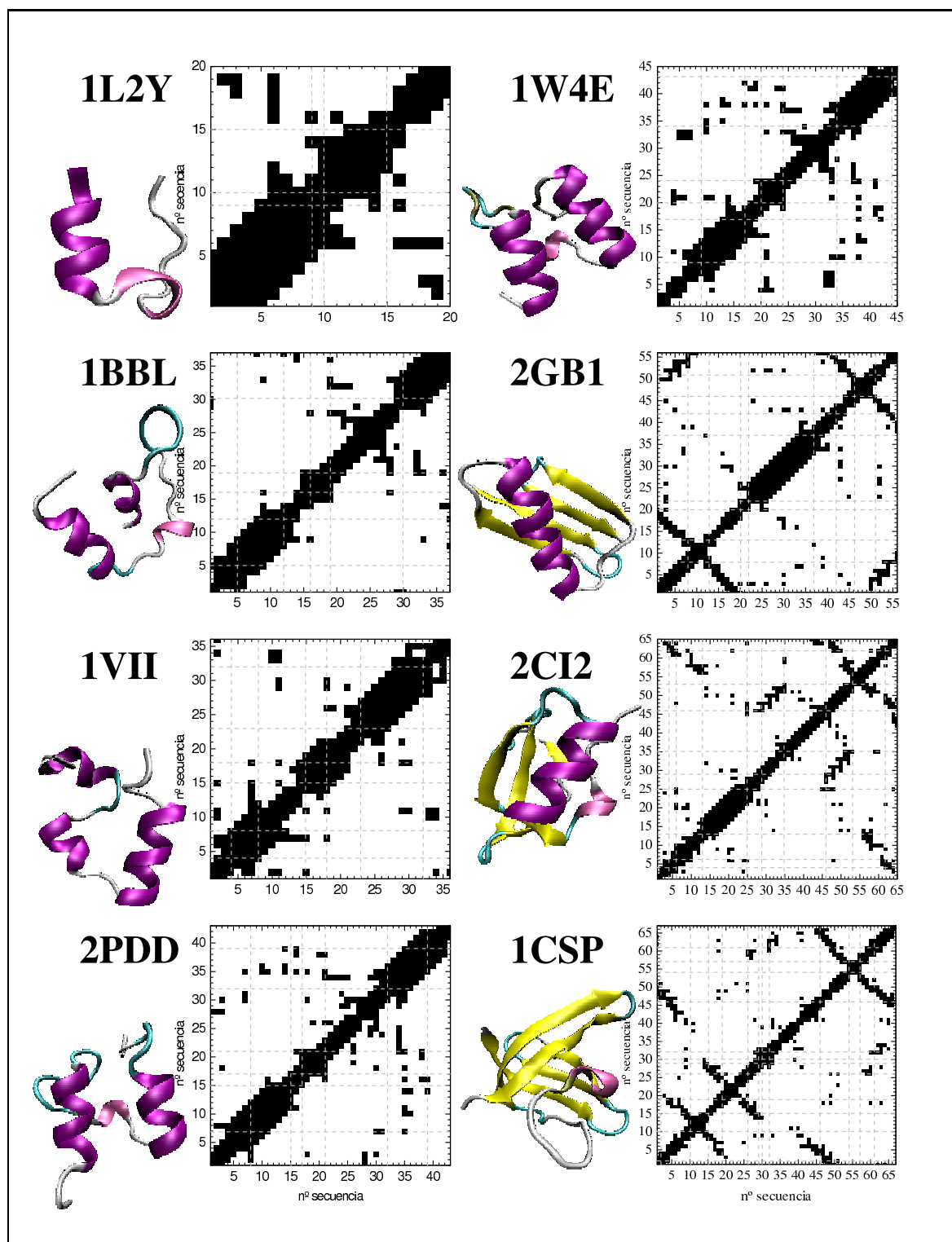


Figura 4.8: Conjunto de proteínas estudiadas en esta sección. La estructura tridimensional (dibujada con VMD⁶⁹) aparece a la izquierda del mapa de contactos correspondiente. Las líneas discontinuas en los mapas separan los elementos de estructura secundaria.

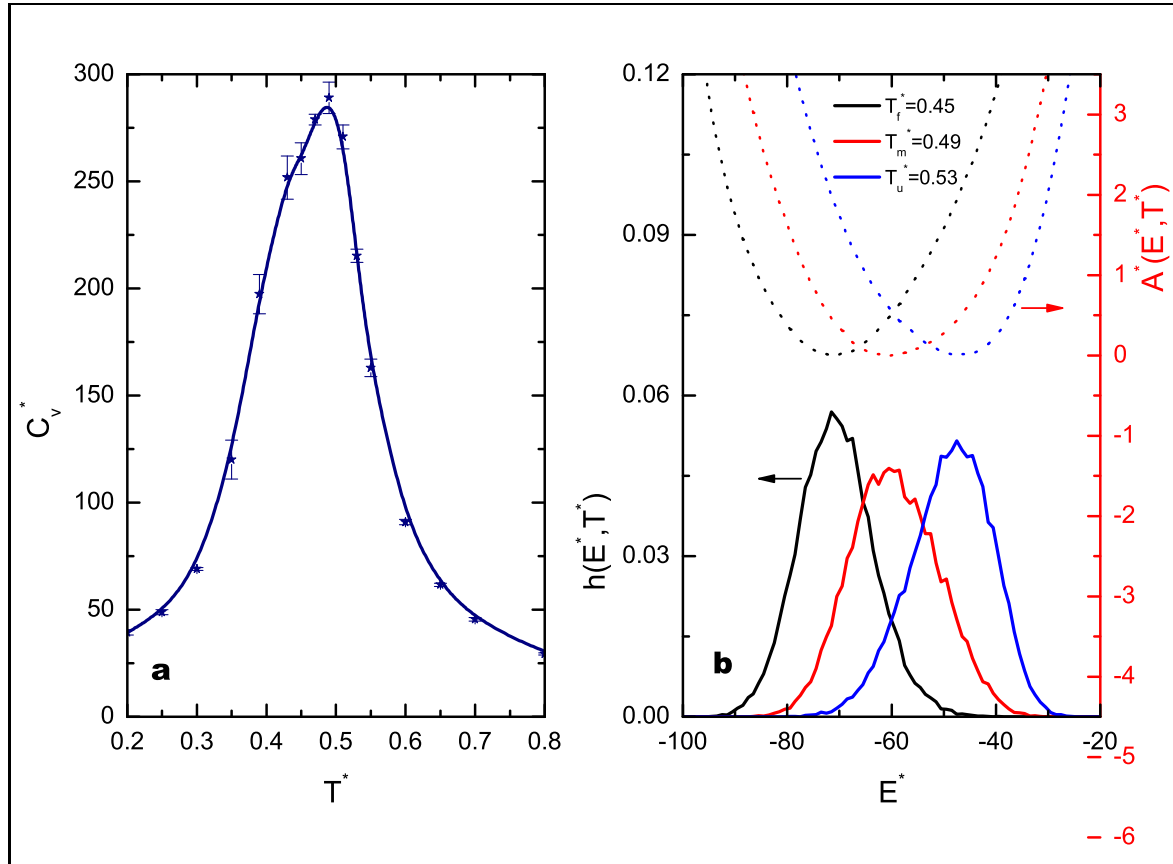


Figura 4.9: Propiedades termodinámicas de la transición de plegamiento de 1BBL: **a)** variación de la capacidad calorífica (C_v^*) frente a la temperatura; **b)** histogramas de energía (líneas continuas) y perfiles de energía libre (líneas punteadas) para la temperatura de transición (T_m^*) y para $T_f^* < T_m^*$ y $T_u^* > T_m^*$.

distribución de energía, así como los perfiles de energía libre, a diferentes temperaturas (Figura 4.9b). La curva de capacidad calorífica de esta proteína presenta un único pico. Vemos que el valor del máximo de C_v^* obtenido para esta proteína es mucho menor que para 2GB1 (con $w^l = w^{lr}$, ver Figura 4.1). Esto ya nos da una idea de que, a esa temperatura, las fluctuaciones de energía son menores. Además, la anchura del pico es mucho mayor que en el caso de 2GB1. En 1BBL, el pico tiene una anchura a mitad de altura de un 40 % con respecto a T_m^* , mientras que para 2GB1 esta anchura es del 4 %. La transición en el caso de 1BBL es mucho menos abrupta que la de 2GB1. De hecho, como indican el perfil de energía libre con un único mínimo y el histograma de energía unimodal a T_m^* (incluidos en la Figura 4.9b), la transición de plegamiento no se produce

de manera cooperativa. A medida que disminuye la temperatura, el mínimo de energía libre se desplaza a valores menores de energía. Pero no aparecen en ningún momento dos estados simultáneamente, sino que la transformación del estado desnaturalizado al plegado es gradual.

¿Se puede decir entonces que todas aquellas proteínas que tienen un porcentaje bajo de contactos terciarios en el estado nativo presentan un plegamiento sin barrera? Si esto es así, ¿cuál es el valor límite para determinar el tipo de transición del plegamiento de una proteína?

Las proteínas de la Figura 4.8 tienen diferentes contribuciones de energía no local en el estado nativo. Dichas contribuciones se encuentran en la Tabla 4.3. Las tres proteínas más grandes del conjunto, 2GB1, 2CI2 y 1CSP, tienen porcentajes de energía de largo alcance dentro del intervalo mayoritario. Por lo tanto, es de esperar que su plegamiento responda a un modelo de dos estados. Por otra parte, las proteínas más pequeñas, 1L2Y y 1BBL, con un 36 % de contactos de largo alcance en el estado nativo, claramente están fuera de dicho intervalo y, como hemos visto para 1BBL, es de esperar que se plieguen a través de un proceso sin barrera. En el caso de 1VII, 2PDD y 1W4E ya no es tan fácil determinar *a priori* cuál va a ser su comportamiento. En la Figura 4.7 vemos que la mayoría de proteínas están entre un 50 y un 60 % de contribución de largo alcance, pero ¿podemos decir que fuera de ese intervalo todas las proteínas presentan un plegamiento sin barreras?

A partir de los resultados de las simulaciones realizadas para todas las proteínas del conjunto propuesto podemos calcular el perfil de energía libre de cada una de ellas a la temperatura de transición correspondiente, determinada como aquella a la que aparece un máximo en la capacidad calorífica. Estas temperaturas aparecen también en la Tabla 4.3. Los perfiles de energía libre nos permiten determinar cuáles son las características termodinámicas de la transición que tiene lugar en nuestras simulaciones para cada una de las proteínas estudiadas.

Tabla 4.3: Datos estructurales y temperatura de transición (T_m^*) para las distintas proteínas estudiadas. $n^o aa$ es el número de residuos de cada proteína; $\%E_{lr}^*$ es el porcentaje de energía debida a interacciones de largo alcance en el estado nativo; N_{con} es el número total de contactos y N_N^Z y N_N^P indican el número de contactos de largo alcance por residuo en el estado nativo según la clasificación de los contactos establecida por Zuo y colaboradores⁹⁵ y según el criterio establecido aquí, respectivamente.

| PDB | 1L2Y | 1BBL | 1VII | 2PDD | 1W4E | 2GB1 | 2CI2 | 1CSP |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|
| $n^o aa$ | 20 | 37 | 36 | 43 | 45 | 56 | 65 | 67 |
| $\%E_{lr}^*$ | 36 | 36 | 43 | 42 | 47 | 50 | 51 | 52 |
| N_{con} | 55 | 108 | 117 | 140 | 161 | 215 | 257 | 270 |
| T_m^* | 0.480 | 0.490 | 0.525 | 0.535 | 0.625 | 0.630 | 0.670 | 0.683 |
| N_N^Z | 0.40 | 0.73 | 0.67 | 0.88 | 1.07 | 1.79 | 1.86 | 2.13 |
| N_N^P | 1.00 | 1.05 | 1.39 | 1.37 | 1.67 | 1.93 | 2.02 | 2.10 |

Los perfiles de energía libre obtenidos para cada proteína a su correspondiente temperatura de transición están representados en la Figura 4.10. Efectivamente, los perfiles de energía libre de 2GB1, 2CI2 y 1CSP corresponden a una transición de primer orden, dada la presencia de una barrera importante de energía libre. Por otra parte, como habíamos predicho, 1LY2 y 1BBL, debido a la baja contribución de las interacciones de largo alcance a la estabilidad del estado nativo, presentan un plegamiento sin barrera. Estos resultados coinciden, en líneas generales, con las observaciones experimentales comentadas al principio de esta sección.

Las proteínas cuya contribución de las interacciones terciarias a la energía del estado nativo no dejan claro qué tipo de transición tiene lugar, como son 1VII, 2PDD y 1W4E, presentan comportamientos diferentes. Por un lado, 1VII y 2PDD presentan un perfil de energía libre que corresponde a una transición de tipo *downhill*. En el caso de 1VII, si bien es cierto que aparece una barrera, ésta es muy pequeña y consideramos que no es lo suficientemente significativa como para considerar que su plegamiento co-

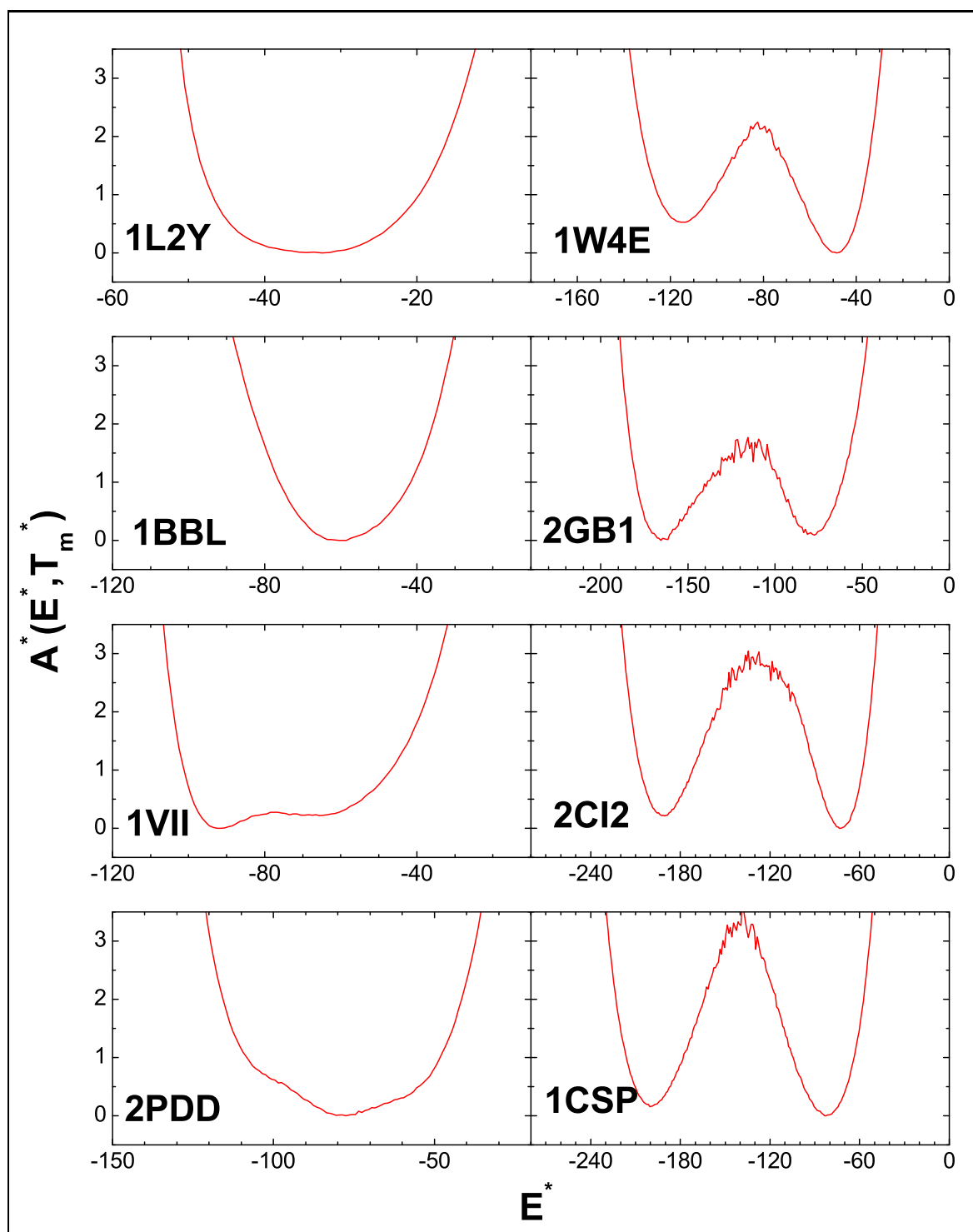


Figura 4.10: Perfil de energía libre para las proteínas de la Figura 4.8 a T_m^* .

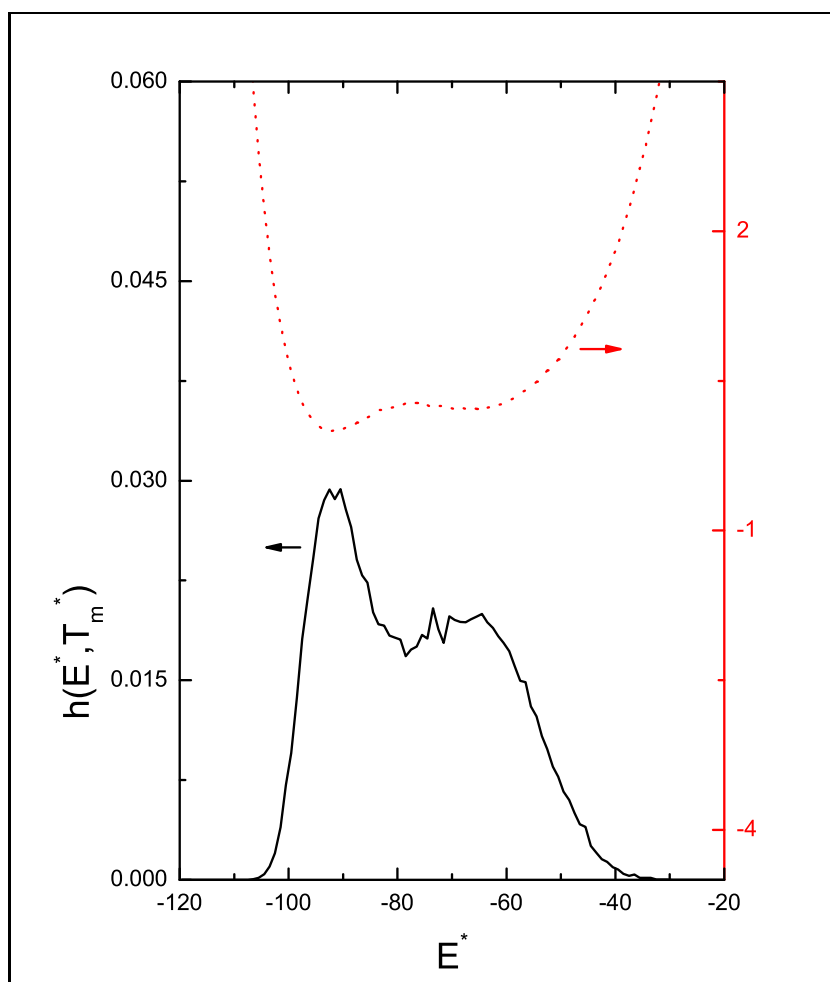


Figura 4.11: Perfil de energía libre (línea roja punteada) e histograma de energía (línea negra continua) a T_m^* para la transición de 1VII.

responde a un modelo de dos estados. De hecho, hay estudios experimentales según los cuales, aunque 1VII cruza una barrera de energía libre al plegarse, ésta es muy pequeña¹⁴⁸, de ahí su elevada velocidad de plegamiento. A partir de los resultados de nuestras simulaciones, ampliados en la Figura 4.11, podemos comprobar que la barrera de energía libre que aparece no es lo suficientemente grande como para que la población de estados con energía intermedia entre la de los estados plegado y desnaturalizado pueda considerarse nula. Vemos que el histograma presenta un pico y un hombro en lugar de dos picos bien diferenciados. Por lo tanto, podemos decir que el plegamiento de 1VII es prácticamente una transición gradual.

Al contrario que 1VII y 2PDD, 1W4E presenta, según nuestro modelo, un plegamiento de dos estados perfectamente definido (ver Figura 4.10). El porcentaje de energía no local en su estado nativo es de un 47 %. Este valor se aleja del 50 % menos que el correspondiente a 1VII y 2PDD (43 y 42 %, respectivamente).

En resumen, se puede decir que la topología del estado nativo tiene una gran influencia sobre las propiedades termodinámicas de la transición de plegamiento, al menos en proteínas pequeñas. Podemos entonces, utilizando un potencial basado en la topología, explicar por qué ciertas proteínas podrían presentar un plegamiento tipo *downhill*, aunque deben ser los resultados experimentales los que, en última instancia, determinen cuál es el tipo de transición que tiene lugar.

4.4. Transición termodinámica, topología y rigidez de cadena.

En la Sección 4.2 hemos visto que las propiedades de la transición varían si se modifica la contribución de las interacciones de largo y corto alcance a la energía del estado nativo. En dicha sección, utilizando 2GB1, hemos visto que un porcentaje entre un 25 y un 63 % de energía terciaria en el estado nativo da lugar a una transición de dos estados.

Las proteínas del apartado anterior que pliegan a través de una transición sin barrera, 1L2Y, 1BBL, 1VII y 2PDD, presentan un porcentaje de contactos de largo alcance respecto del total de 36, 36, 43 y 42 %, respectivamente. Todos estos valores entran dentro del intervalo en que la transición de 2GB1 está correctamente descrita, pero en el caso de 1L2Y y 1BBL, por ejemplo, es un valor bastante bajo. Además, ya hemos comentado que ese intervalo no tiene por qué ser generalizable a todas las proteínas. Para comprobar que los resultados obtenidos para estas proteínas no son un artificio del potencial y que con un reajuste apropiado de los parámetros no obtenemos

otro tipo de transición, hemos calculado los perfiles de energía libre que se obtienen para estas proteínas variando las contribuciones a la energía del estado nativo. Estos perfiles, calculados como siempre a T_m^* , aparecen en la Figura 4.12. En el caso de 1LY2 y 1BBL, variando los valores de w^l y w^{lr} hemos aumentado la contribución de las interacciones terciarias a, aproximadamente, un 50 y un 70 %. Para 1VII y 2PDD, que sin variar los valores de w^l y w^{lr} tienen una contribución cercana al 50 %, por un lado hemos disminuido esta contribución y por otro la hemos aumentado. Aunque se pueden observar ligeras variaciones de unas contribuciones a otras, como el valor de energía a la que aparece el mínimo de energía libre, las propiedades de la transición no cambian de manera sustancial. Estas proteínas, que sin modificar los valores naturales de w^l y w^{lr} presentan un plegamiento sin barrera, mantienen este comportamiento incluso para contribuciones elevadas de las interacciones no locales ($\simeq 70\%$).

Para las proteínas que con este modelo presentan un plegamiento de dos estados, la rigidez de cadena modifica las propiedades termodinámicas de manera similar a lo observado para 2GB1 (ver Figura 4.13). Vemos que se produce una variación en la altura de la barrera de energía libre y una variación en ΔE_m^* de unas contribuciones a otras. En cualquier caso, para casi todas las contribuciones utilizadas, la transición se mantiene como una transición de dos estados. Por lo tanto, aunque modificando los parámetros y haciendo que las contribuciones de un tipo de interacción sea extrema se puede hacer desaparecer la barrera de energía libre, el plegamiento mantiene sus propiedades fundamentales ante las variaciones en la contribución de los dos tipos de interacción que estamos considerando.

En resumen, podemos decir que, con el potencial que estamos utilizando, es la topología del estado nativo en sí misma la que determina que la transición de plegamiento transcurra a través de uno o dos estados, sin o con barreras.

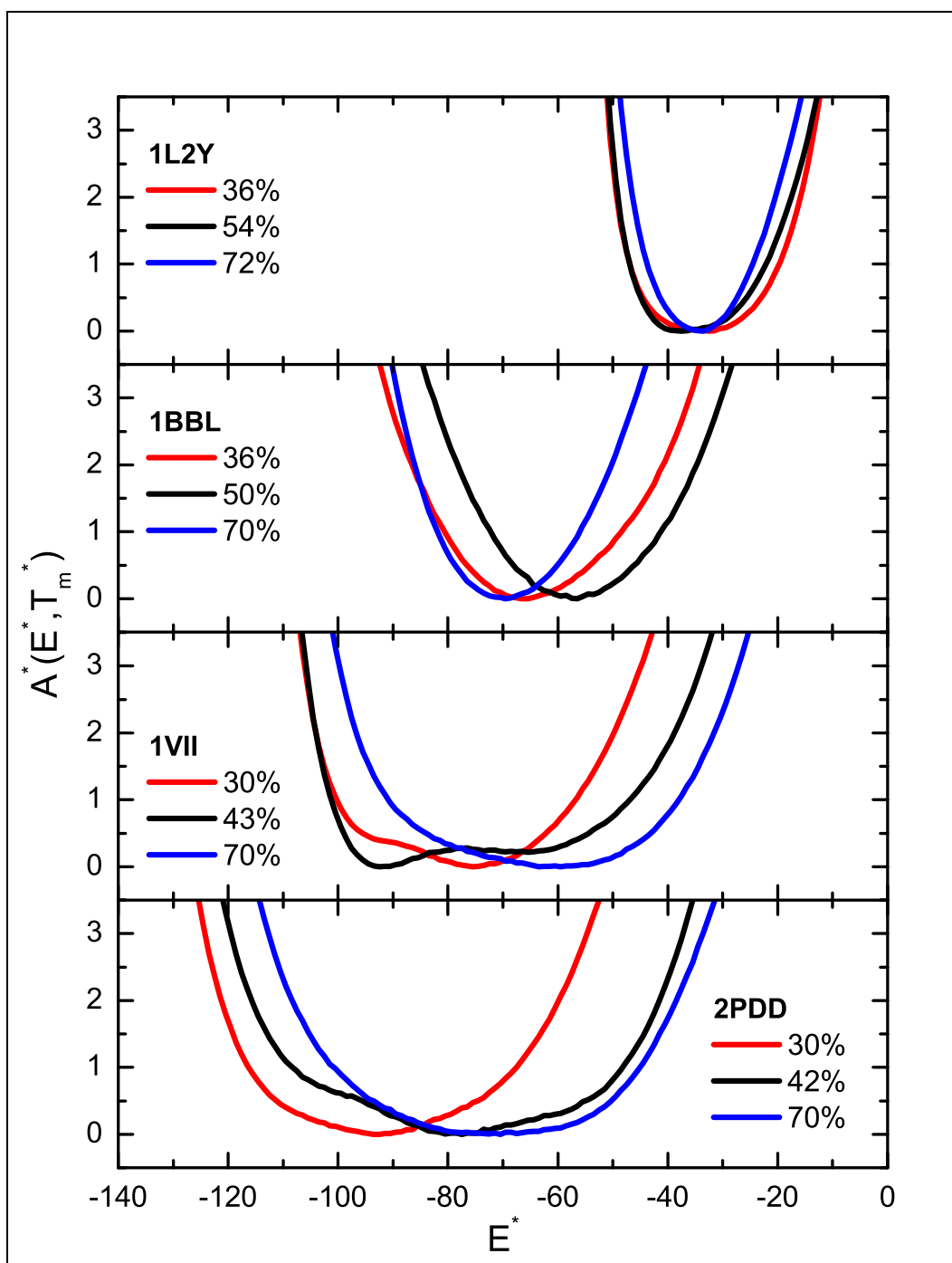


Figura 4.12: Perfiles de energía libre a la temperatura de transición para proteínas con un plegamiento sin barrera según nuestro modelo, obtenidos a partir de las distintas contribuciones de las interacciones no locales a la estabilidad del estado nativo.

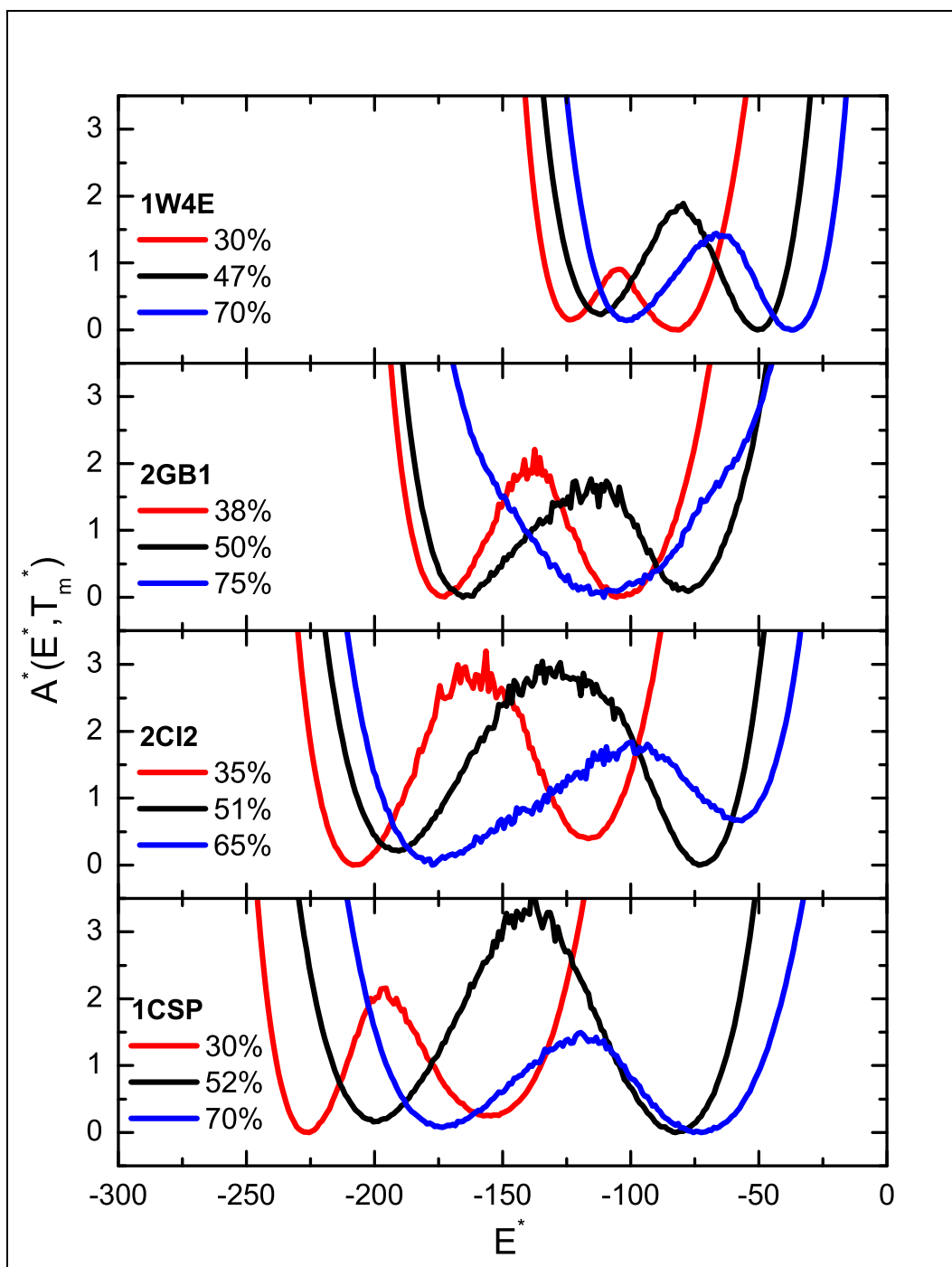


Figura 4.13: Perfiles de energía libre a la temperatura de transición para proteínas con un plegamiento de dos estados según nuestro modelo, obtenidos a partir de las distintas contribuciones de las interacciones no locales a la estabilidad del estado nativo.

4.5. Análisis de la topología del estado nativo.

¿Cómo determinar *a priori* si una proteína con una topología determinada da lugar a un tipo de transición de plegamiento u otro? ¿Qué características concretas de la topología dan esa información?

Hemos visto que la mayoría de las proteínas presentan entre un 50 y un 60 % de contactos terciarios en el estado nativo y que estas proteínas presentan un plegamiento de dos estados. Además, ciertas proteínas pequeñas tienen una proporción de contactos de largo alcance por debajo del 50 %. Algunas de estas proteínas, cuya proporción de contactos terciarios no se aleja demasiado del 50 %, presentan también una transición de primer orden. Sin embargo, aquéllas que tienen una proporción de estos contactos significativamente baja se pliegan a través de una transición de tipo *downhill*. No es fácil determinar, a partir de esta propiedad, qué proteínas con menos de un 50 % de estructura terciaria en el estado nativo presentan un plegamiento con o sin barrera.

En un trabajo reciente⁹⁵, Zuo y colaboradores, utilizando un modelo similar al que utilizamos en este trabajo y realizando simulaciones de dinámica de Langevin, han establecido que existe una correlación entre la cooperatividad y el número de contactos terciarios por residuo:

$$N_N = \frac{\text{n}^\circ \text{ contactos terciarios}}{\text{n}^\circ \text{ aminoácidos}}. \quad (4.2)$$

De hecho, definen un valor límite de $N_N = 0.9$ por encima del cual se puede decir que la proteína presenta un plegamiento de dos estados. El valor concreto de N_N depende de la clasificación de contactos que se esté utilizando. En el trabajo de Zuo y colaboradores, los contactos entre un aminoácido i y el $i+4$ están incluidos dentro del conjunto de contactos locales. En la Tabla 4.3 aparecen los valores de N_N calculados según la clasificación de contactos terciarios de Zuo (N_N^Z) y según nuestra clasificación (N_N^P). Aunque los valores son distintos según la clasificación de contactos utilizada, la tendencia general es la misma en los dos casos. Además vemos que, efectivamente, el valor $N_N^Z = 0.9$ separa

las proteínas que pliegan según una transición de dos estados de aquéllas que lo hacen siguiendo un proceso sin barrera. Queda establecida así una clara relación entre el tipo de transición y la contribución de los contactos de largo alcance a la estabilidad del estado nativo. Con nuestra clasificación de las interacciones, el valor límite de N_N^P se encontraría entre 1.4 y 1.5, dejando clara la validez del tratamiento de Zuo y colaboradores.

Sin embargo, el valor de parámetros como la longitud de secuencia o la proporción de los contactos de largo alcance siguen sin dar respuesta a la pregunta planteada: ¿Qué características concretas de la topología determinan las propiedades termodinámicas del plegamiento de una proteína? Volviendo a la Figura 4.8, podemos comprobar que, a primera vista, las proteínas de la izquierda, que son las que presentan un plegamiento sin barrera, tienen un mapa de contactos con una estructura terciaria peor definida que aquéllas que pliegan siguiendo un proceso de dos estados (las de la derecha en la misma figura). Esta característica está reflejada, en gran medida, en el parámetro N_N . Pero, ¿hasta qué punto la estructura terciaria es clave en el tipo de transición?

Si nos fijamos en los mapas de contactos de 2PDD y 1W4E, podemos ver que son muy parecidos. Esto es razonable, teniendo en cuenta que 1W4E es un mutante de 2PDD. Sin embargo, las transiciones a que dan lugar son muy diferentes: 1W4E presenta un plegamiento de dos estados mientras que 2PDD lo hace a través de un mecanismo sin barrera. Las principales diferencias estructurales entre una y otra proteína están en su longitud (1W4E tiene dos aminoácidos más) y en el número de contactos. 2PDD tiene 81 contactos locales y 59 de largo alcance. 1W4E tiene 85 contactos locales, cuatro más que 2PDD debido a su mayor longitud, y 76 contactos terciarios. ¿Cómo varía la topología de 1W4E con los 17 contactos extra que tiene con respecto a 2PDD? En los mapas de contactos de la Figura 4.8 no podemos apreciar, a simple vista, diferencias sustanciales en la estructura terciaria. Aunque se aprecia que hay ligeras diferencias, la topología parece la misma. En el Capítulo 3 hemos visto que pequeñas diferencias en los mapas de contactos afectan a propiedades termodinámicas como la altura de la barrera o la

entropía del estado plegado, aunque no son lo suficientemente importantes como para dar lugar a transiciones de diferente tipo. Sin embargo, en el caso de 2PDD y 1W4E, topologías muy similares dan lugar a transiciones diferentes.

Dado que el mapa de contactos a simple vista no da información suficiente para determinar las diferencias entre las topologías de 2PDD y 1W4E, hemos realizado un análisis más detallado de las mismas. En la Figura 4.14 hemos calculado el histograma de la distancia que separa a lo largo de la secuencia a dos aminoácidos que están en contacto. Hemos calculado cuántos pares que forman un contacto en el estado nativo están a una distancia $|i - j|$ a lo largo de la secuencia, es decir, hay $|i - j| - 1$ aminoácidos entre ellos (siendo i y j el número de residuo a lo largo de la secuencia de aquellos aminoácidos que están en contacto). Una vez hecho esto, para poder comparar unas estructuras con otras, hemos dividido esta cantidad entre el número total de contactos de largo alcance. En resumen, estamos representando el histograma de órdenes de contacto absolutos del estado nativo⁴⁰. En estas gráficas obtenemos información adicional sobre las diferencias estructurales entre las proteínas que dan lugar a un tipo de transición u otro. En la mayoría de los casos, estos histogramas presentan un pico para un valor de $|i - j| = 4$. Este valor corresponde a los contactos característicos de las hélices α . Aunque nosotros hemos incluido estos contactos dentro del conjunto de contactos de largo alcance, es innegable su carácter local, dado que estabilizan elementos de estructura secundaria (hélices α). Esto sigue confirmando el hecho de que un grado demasiado elevado de estructura local (helicoidal en este caso) disminuye la cooperatividad de la transición de plegamiento. El pico correspondiente a los contactos $|i - j| = 4$ tiene un valor por encima de 0.3 para todos los casos de proteínas con un plegamiento sin barrera. Esto no quiere decir, sin embargo, que la presencia de hélices implique directamente una transición no cooperativa, aunque en otros trabajos se ha postulado la posibilidad de que sean las responsables de la presencia de barreras bajas e, incluso, despreciables¹⁴⁹. Muchas proteínas tienen estructura helicoidal bien definida, pero es la presencia de una

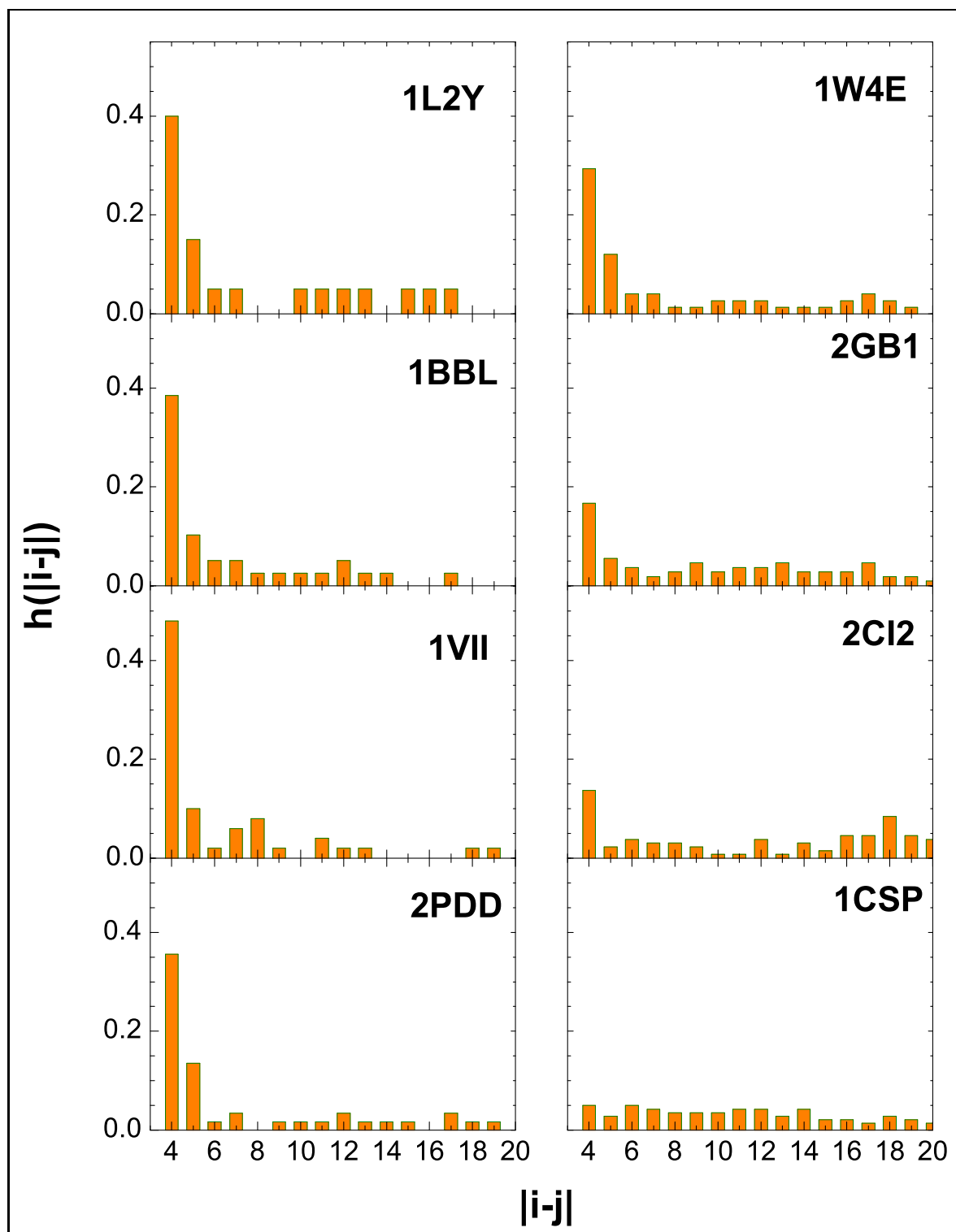


Figura 4.14: Histogramas de órdenes de contacto nativos para todas las proteínas de la Figura 4.8. i y j son los números de residuo a lo largo de la secuencia de los aminoácidos que forman un contacto. Los histogramas están referidos al número total de contactos de largo alcance que presenta cada proteína.

estructura terciaria lo suficientemente compleja como para que exista una compensación entre las interacciones locales y las de largo alcance, que como hemos visto se traduce en un equilibrio entre energía y entropía, lo que da lugar a una barrera de energía libre.

La diferencia entre los histogramas de orden de contacto de unas proteínas y otras no es sólo la proporción de contactos helicoidales. Podemos observar también que las proteínas que tienen un plegamiento de dos estados presentan una distribución más homogénea de los diferentes valores de orden de contacto. Ésta es la diferencia fundamental entre los histogramas de orden de contacto de 2PDD y 1W4E. Ambas proteínas presentan una proporción de contactos helicoidales muy parecida, siendo algo menor la de 1W4E. La distribución del resto de los contactos también es muy similar en ambos casos. Sin embargo, los contactos de largo alcance que 1W4E tiene de más con respecto a 2PDD corresponden a contactos con $|i - j| > 4$. Esto hace que la proporción de contactos helicoidales sea menor en 1W4E y, además, que el histograma de orden de contacto sea más homogéneo. Todos los valores de $|i - j|$ están poblados en el caso de 1W4E, mientras que 2PDD presenta huecos libres.

Hasta ahora hemos visto que, para que la transición de plegamiento sea de primer orden, es necesario un equilibrio entre los contactos de largo y corto alcance. A partir de los resultados de la Figura 4.14 podemos concluir que, además, también es necesario un equilibrio dentro del conjunto de contactos terciarios. Dentro de ellos, según nuestra definición, hay una proporción importante de contactos entre aminoácidos cercanos en secuencia. Estos contactos son los máximos responsables de la estructura en el caso de proteínas cuyo plegamiento transcurre sin cruzar una barrera de energía libre. Estas proteínas presentan también contactos con valores mayores de $|i - j|$, pero en baja proporción, lo cual hace la distribución del histograma $h(|i - j|)$ poco homogénea, presentando huecos para valores de $|i - j| > 4$. Por otro lado, las proteínas con un plegamiento de dos estados no presentan huecos en dicho histograma, siendo la distribución a lo largo de todos los órdenes de contacto (hasta cierto valor de $|i - j|$) bastante

equitativa. Ésta es una diferencia importante entre las proteínas que pliegan de uno u otro modo y es lo que, principalmente, diferencia la topología de 2PDD y 1W4E. Por otra parte, la distribución homogénea de los contactos a lo largo de todos los valores de $|i - j|$ es también en cierto modo indicativa de que los contactos están distribuidos entre todos los elementos de estructura secundaria. Como hemos visto en el análisis de las transiciones de 2GB1, 1PGB y sus “mutantes”, esto es de gran importancia para la cooperatividad de la transición de plegamiento.

Con objeto de comprobar la generalidad de esta observación, hemos realizado el mismo análisis de órdenes de contacto para todas las proteínas de la base de datos utilizada anteriormente. Hemos calculado el histograma de orden de contacto de las 1590 proteínas referido al número de contactos de largo alcance que presenta cada una. La media de esos histogramas son los que podemos ver en la Figura 4.15. Hemos separado las proteínas más pequeñas (igual o menos de 60 residuos) de las más grandes para evitar que los resultados de éstas (que son muchas más) enmascaren los resultados de las primeras. A pesar de que el promediado oculta las características concretas de cada proteína, se pueden sacar conclusiones generales: por un lado, la proporción de estructura helicoidal en las proteínas pequeñas es mucho mayor que en las grandes; en segundo lugar, la distribución del orden de contacto es bastante homogénea a lo largo de todo el intervalo de valores de $|i - j|$ representado; además, las barras de error de los histogramas son muy pequeñas, lo cual indica que las conclusiones que se puedan sacar de estos resultados son bastante generalizables. Dado que la mayoría de las proteínas estudiadas hasta el momento parece que presentan un plegamiento experimental de dos estados, los resultados presentados en la Figura 4.15 podrían ser indicativos de características estructurales que dan lugar a este tipo de transición. Los casos de proteínas pequeñas que presentan una proporción de contactos locales elevada y una distribución inhomogénea de los órdenes de contacto presentan, según nuestro modelo, un plegamiento sin barrera, como es el caso de 1L2Y, 1BBL, 1VII y 2PDD,

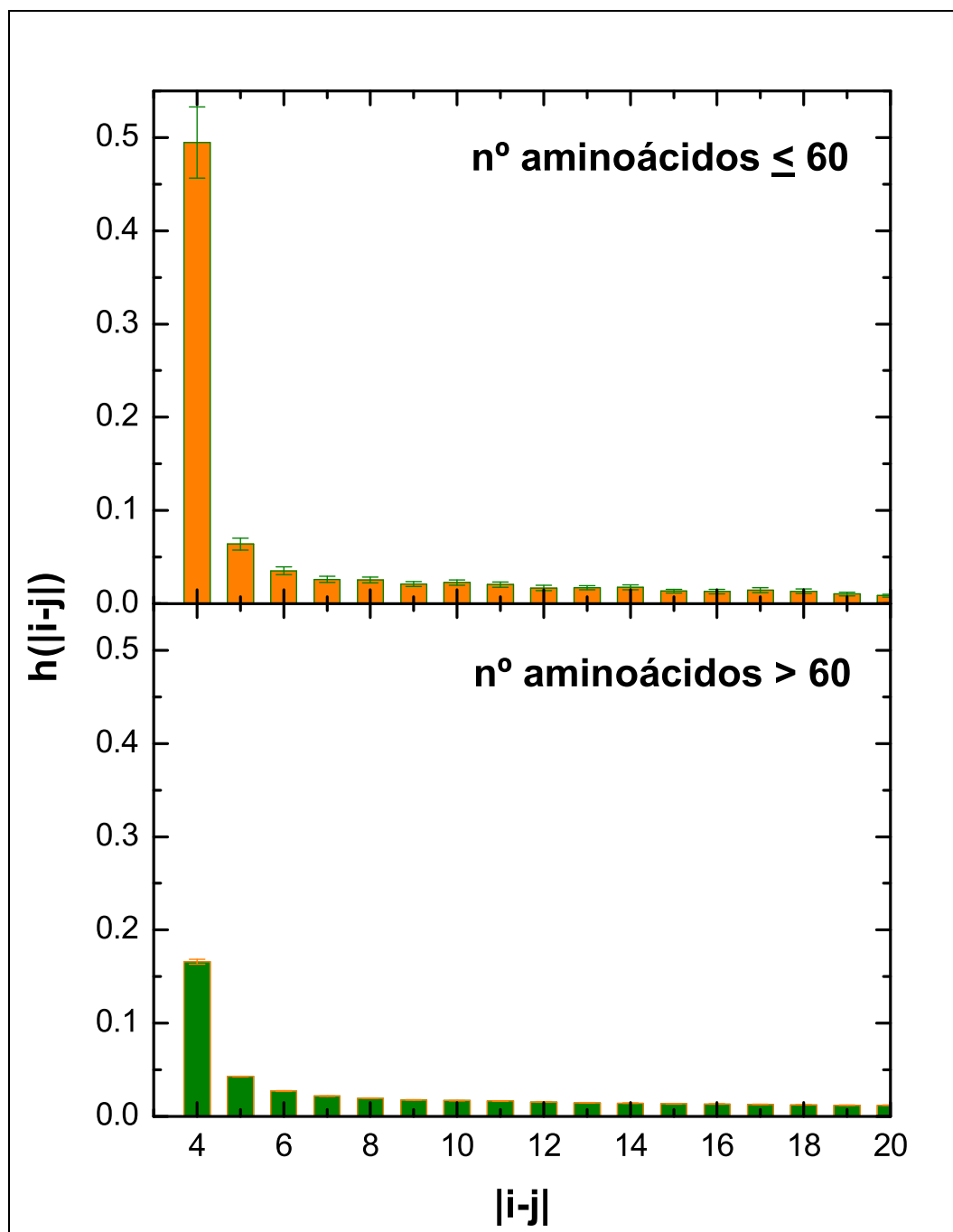


Figura 4.15: Promedio de los histogramas de orden de contacto de una base de datos de 1590 proteínas.

lo cual podría explicar las observaciones experimentales realizadas a partir de estudios experimentales sobre estas mismas proteínas.

4.6. Mutaciones del mapa de contactos.

Al final del Capítulo 3 hemos estudiado cómo varían las propiedades de la transición con pequeñas modificaciones en los mapas de contactos. Hemos podido ver que, aunque las características generales de los dos mapas de contactos comparados determinan que la transición sea de primer orden, las propiedades de la transición varían ligeramente debido a pequeñas diferencias en los mapas. En este capítulo hemos visto además que mapas de contactos con bastante similitud, como es el caso de 2PDD y 1W4E, dan lugar a transiciones radicalmente diferentes. En la sección anterior hemos encontrado diferencias en la topología de estas dos proteínas que podrían explicar las diferentes características de la termodinámica de la transición que presenta cada una de ellas.

Para ahondar más en la trascendencia de estas diferencias, en esta sección estudiamos específicamente cómo afectan las modificaciones del mapa de contactos a las transiciones de plegamiento de 2PDD y 1W4E. Así podemos profundizar en las características concretas de la topología que determinan a su vez las características termodinámicas de sus respectivos plegamientos.

Para poder comparar los mapas de las proteínas es necesario, primero, ver qué aminoácidos de una proteína son equivalentes a los de la otra. Quitando los dos primeros aminoácidos de 1W4E y el último de 2PDD, de manera que reducimos ambas proteínas a 42 aminoácidos, resulta que las secuencias de estas dos proteínas sólo se diferencian en un residuo. Este residuo es el 38, que es fenilalanina en 2PDD y triptófano en 1W4E (ver Figura 1.2). 1W4E es un mutante de 2PDD y es de esperar que la estructura de ambos sea muy parecida. El valor de *rmsd* entre las dos estructuras es de 3.1 Å (medido a partir de las posiciones de los C_α), que no es un valor demasiado elevado, aunque

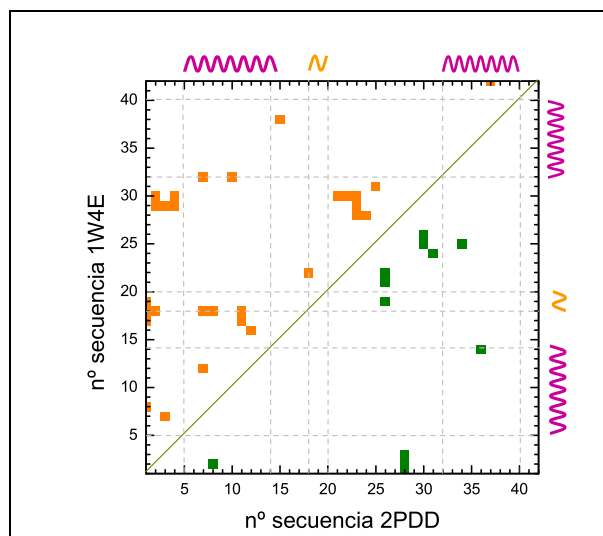


Figura 4.16: Mapa de diferencias entre los mapas de 1W4E y 2PDD. En el triángulo superior, en naranja, están los contactos que tiene 1W4E que no aparecen en el mapa de 2PDD; en el triángulo inferior, en verde, los que tiene 2PDD que no aparecen en el mapa de 1W4E. Las hélices marcadas en rosa son comunes a ambas proteínas. La pequeña hélice de color naranja sólo aparece en 1W4E.

tampoco despreciable. Aunque a grandes rasgos ambas estructuras se parecen mucho, hemos visto que las diferencias en sus topologías son lo suficientemente importantes como para dar lugar a una transición diferente. Por lo tanto, es necesario determinar cuáles son esas diferencias de manera más detallada. Con este objetivo hemos representado qué contactos son los que aparecen en un mapa y no en el otro, como ya hemos hecho en el capítulo anterior. Esta representación aparece en la Figura 4.16, donde podemos apreciar diferencias significativas entre las dos topologías. Además, en 1W4E hay una pequeña región helicoidal que en 2PDD no aparece. Los mapas no sólo se diferencian en que 1W4E tenga más contactos, sino que, además, hay 12 contactos que aparecen en 2PDD que no aparecen en 1W4E, y viceversa. Es decir, los contactos terciarios aparecen distribuidos de diferente manera. Como ya hemos visto para el caso de los mapas de 1PGB y 2GB1, algunos contactos, aunque no son iguales en las dos proteínas, sí son equivalentes. Así, podemos decir que son equivalentes los pares de la Tabla 4.4. Estos contactos, además de corresponder a zonas de interacción equivalentes entre ambos

Tabla 4.4: Contactos equivalentes en 1W4E y 2PDD.

| 1W4E | | | | 2PDD | | |
|------|-----|-----------|--|------|-----|-----------|
| i | j | $ i - j $ | | i | j | $ i - j $ |
| 1 | 8 | 7 | | 2 | 8 | 6 |
| 2 | 29 | 27 | | 1 | 28 | 27 |
| 3 | 29 | 26 | | 2 | 28 | 26 |
| 4 | 29 | 25 | | 3 | 28 | 25 |
| 23 | 28 | 5 | | 25 | 30 | 5 |
| 24 | 28 | 4 | | 26 | 30 | 4 |
| 23 | 30 | 7 | | 24 | 31 | 7 |
| 15 | 38 | 23 | | 14 | 36 | 22 |

mapas, tienen un orden de contacto idéntico, excepto en dos casos en los que es muy similar. Por lo tanto, es razonable pensar que son los otros 21 contactos diferentes entre 1W4E y 2PDD los que hacen que la transición de 1W4E presente una barrera de energía libre que no encontramos en 2PDD. Para comprobar cómo afectan a la transición estos contactos, hemos creado diferentes “mutantes”, de la misma forma que lo hacíamos en la Sección 3.5. En cada uno de estos mutantes hemos buscado poder estudiar la validez del parámetro N_N o la de los histogramas de orden de contacto.

Hemos generado cuatro mapas de contactos diferentes para 1W4E. Dichos mapas surgen desactivando contactos del mapa de 1W4E. Su topología y la transición termodinámica a la que dan lugar aparecen en la Figura 4.17:

- **Mapa 1:** En primer lugar hemos “apagado” aquellos contactos de 1W4E que tienen un orden de contacto muy elevado con respecto a los que presenta 2PDD. Hemos eliminado los contactos (2-30), (4-30), (7-32) y (10-32), que presentan un orden de contacto $|i - j|$ igual a 28, 26, 25 y 22, respectivamente. Además de reducir el orden de contacto total de la estructura, disminuye también el valor de

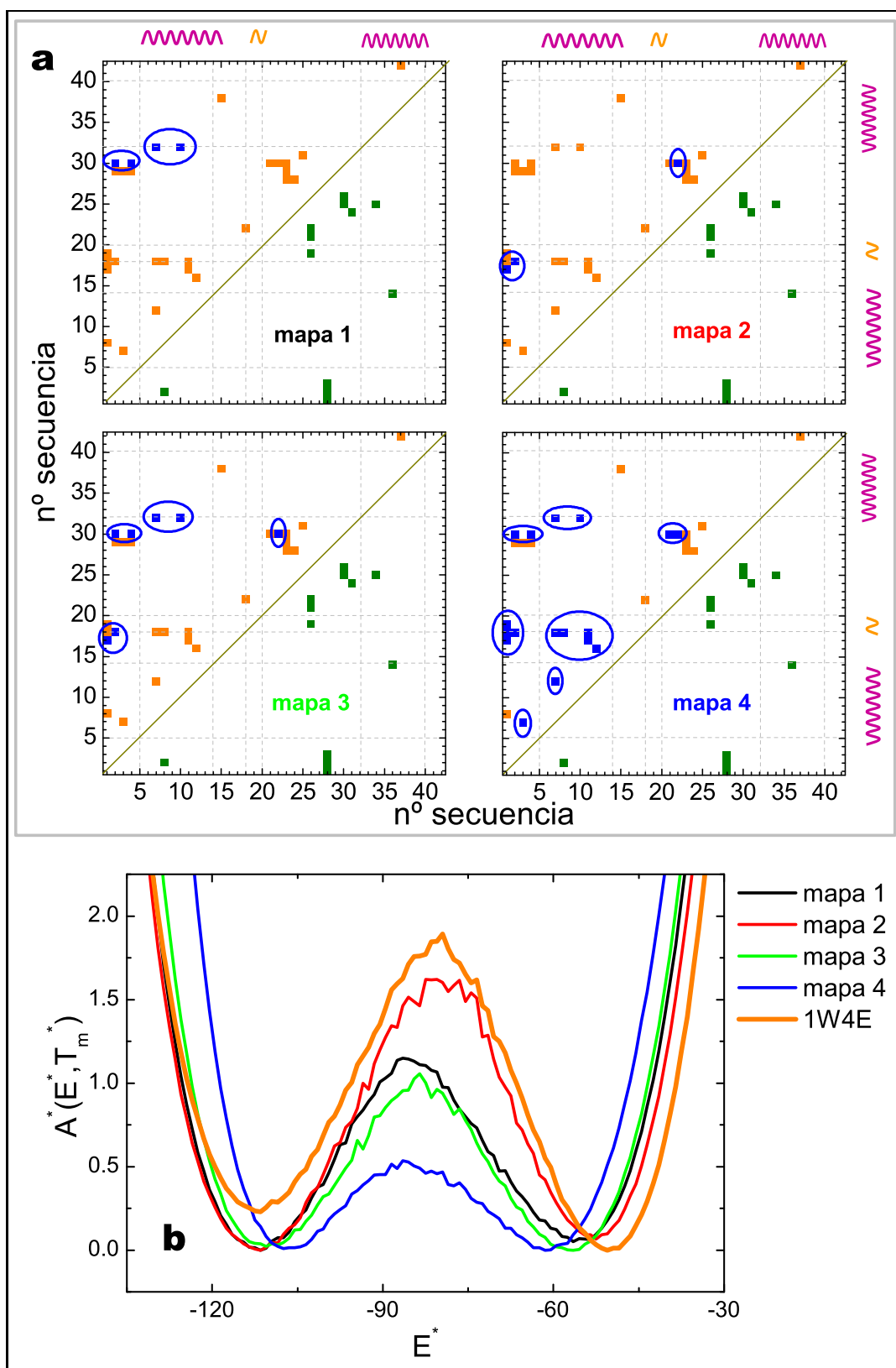


Figura 4.17: **a)** En cada mapa se representan: en verde, los contactos presentes en 2PDD que no están en 1W4E; en naranja, los contactos que tiene 1W4E que no tiene 2PDD y en azul, marcados con un círculo, los contactos que se desactivan en 1W4E. **b)** Perfiles de energía libre de las transiciones a que da lugar cada mapa. La línea naranja corresponde a la transición de 1W4E con el mapa no modificado.

N_N . Según nuestra clasificación de contactos, este parámetro pasaría a tomar un valor $N_N^P = 1.6$, que sigue siendo bastante alto. Vemos en el perfil de energía libre que la barrera disminuye considerablemente con respecto a la de 1W4E sin modificar. Estos contactos corresponden a interacciones entre elementos de estructura secundaria muy alejados en la secuencia, por lo que son cruciales para estabilizar la estructura compacta y, además, para la cooperatividad del tránsito.

- **Mapa 2:** En este caso hemos “apagado” aquellos contactos que presentan un valor de $|i - j|$ igual a 8 y 16. Así, el histograma de orden de contacto de 1W4E pasa a tener vacías las mismas posiciones que el histograma de 2PDD (ver Figura 4.14). Estos contactos son (1-17), (2-18) y (22-30). El parámetro N_N^P tiene, en este caso, un valor de 1.62, muy parecido al del **mapa 1**. Sin embargo, la barrera apenas se modifica con respecto a la original de 1W4E. Podemos decir entonces que, aunque el histograma de orden de contacto nos puede dar información sobre las características generales de la estructura y una idea sobre el tipo de transición que tiene lugar, no sirve para establecer un criterio único para determinar, *a priori*, si el plegamiento se producirá a través de una transición de primer orden o no.
- **Mapa 3:** Combinando los contactos suprimidos en los dos mapas anteriores vemos que, igual que ocurre con el **mapa 2**, eliminar los contactos (1-17), (2-18) y (22-30) no modifica sustancialmente la transición. Los contactos (1-17) y (2-18) corresponden a un conjunto de interacciones entre la hélice pequeña y la primera región no estructurada de la proteína. Eliminar el contacto (22-30) tampoco supone una gran desestabilización de la región a la que pertenece, ya que en ella hay otros contactos que desempeñan la misma función que éste. Por este motivo no son contactos cruciales para la cooperatividad de la proteína.
- **Mapa 4:** Hemos convertido el mapa de contactos de 1W4E en uno equivalente al de 2PDD. Para ello hemos desactivado todos los contactos que no tienen un

equivalente en el mapa de 2PDD. Es decir, en este mapa están los contactos de la Tabla 4.4 y además (18-22), (23-29), (25-31) y (37-42). Estos contactos no tienen un equivalente directo en 2PDD, pero sí puede establecerse cierta similitud con los contactos de 2PDD (19-26), (21-26), (25-34) y (22-26), respectivamente. Los tres primeros pares tienen en común que corresponden a interacciones entre los mismos elementos de estructura secundaria, mientras que (37-42) y (22-26) sólo se parecen en el orden de contacto. En este mapa, el número de contactos de largo alcance es igual que el de 2PDD y el parámetro N_N^P toma un valor de 1.31, menor que el que corresponde a 2PDD. Por otra parte, el perfil de energía libre a T_m^* de la transición a que da lugar este mapa de contactos presenta una barrera de energía libre. Sin embargo, esta barrera es muy pequeña. Es decir, eliminar estos contactos provoca una disminución importante de la barrera de energía libre, de manera que podemos considerar que, con este mapa de contactos, la proteína presenta un plegamiento gradual (la barrera es aproximadamente del mismo orden que la de 1VII, en cuyo caso hemos visto que no es lo suficientemente elevada como para considerar que la transición termodinámica de plegamiento es de primer orden). El parámetro N_N es, por lo tanto, una buena estimación de cuándo una transición de plegamiento se produce a través de un proceso con o sin barrera, a pesar de que, en este caso, podría dar lugar a pensar que la barrera debería desaparecer completamente, dado que es menor que el de 2PDD.

Estos mutantes nos permiten establecer la importancia de la presencia de determinados contactos para la cooperatividad del tránsito. A medida que modificamos el mapa de contactos de 1W4E de la manera comentada anteriormente lo que sucede es que algunas zonas de interacción ven disminuida su estabilidad. En el caso del **mapa 1**, la zona de contacto entre la primera hélice con otros elementos estructurales se queda casi sin contactos y la barrera de energía libre disminuye considerablemente. Sin embargo, eliminar contactos como se hace en el **mapa 2** no produce una variación importante en

la barrera, ya que las zonas de interacción a las que corresponden esos contactos mantienen otros que contribuyen a su estabilidad. Por lo tanto, la diferencia fundamental entre el mapa de 2PDD y el de 1W4E es que en el segundo existe una mejor distribución de los contactos entre todos los elementos de estructura secundaria. Esta característica, como ya hemos observado comparando los mapas de 1PGB y 2GB1, es crucial en la cooperatividad del plegamiento.

Podemos concluir entonces que, aunque N_N y los histogramas de orden de contacto dan una idea de propiedades importantes de la topología del estado nativo, no siempre son suficientes para relacionarla cuantitativamente con la presencia o no de una barrera de energía libre. Pequeñas alteraciones del mapa de contactos pueden provocar grandes variaciones en las características termodinámicas del plegamiento, siempre que afecten a regiones donde otros contactos no compensen el efecto de las modificaciones introducidas.

4.7. Resumen y conclusiones del Capítulo.

En este capítulo hemos estudiado cómo la topología del estado nativo determina las propiedades termodinámicas de la transición de plegamiento. Este estudio nos ha permitido, a la vez, determinar el papel que juegan las distintas interacciones en la presencia de una barrera de energía libre y, por tanto, en la cooperatividad del tránsito.

Inicialmente, hemos estudiado cómo varían las propiedades termodinámicas del plegamiento de 2GB1 con la rigidez de cadena, definida por la intensidad relativa de las interacciones de largo y corto alcance. Hemos podido comprobar cómo la mayor o menor flexibilidad de la cadena tiene implicaciones importantes tanto en las características del tránsito como en la estabilidad de la proteína. Cuando la cadena es extremadamente flexible o extremadamente rígida, aparecen intermedios termodinámicos de plegamiento que no se corresponden con las observaciones experimentales. Además, en estos casos extremos, la transición no presenta las propiedades características de una transición de dos estados, sino que la barrera de energía libre desaparece. Estas observaciones

nos llevan a la conclusión principal de este estudio y es que, para que la transición se produzca a través de un mecanismo de tipo “todo o nada”, debe haber un equilibrio entre las interacciones locales y de largo alcance, ya que ambas tienen un efecto directo sobre la entropía y energía del sistema.

En el caso de cadenas flexibles, es decir, aquellas con una proporción elevada de energía terciaria en su potencial de interacción, tanto el estado plegado como el estado desnaturalizado presentan valores altos de entropía a la temperatura de transición. Por lo tanto, la diferencia en entropía entre los dos estados es baja. Debido a la amplitud de las fluctuaciones conformacionales, las distribuciones de energía de ambos estados solapan, dando lugar a un único pico en el histograma de energía, lo que se traduce en la ausencia de barrera de energía libre. Cuando la contribución de las interacciones locales aumenta, las distribuciones de energía para cada estado se hacen más estrechas, sobre todo en el caso del estado plegado, de forma que se diferencian dos estados a T_m^* entre los que aparece una barrera de energía libre. Si sigue aumentando la proporción de energía local en la energía total, la cadena comienza a hacerse más y más rígida. Como hemos comprobado, esto provoca que el estado desnaturalizado retenga una elevada proporción de estructura secundaria residual incluso a elevada temperatura. Por lo tanto, la energía media del estado desnaturalizado disminuye al aumentar la rigidez y la diferencia de energía entre los dos estados se hace más pequeña, provocando de nuevo la desaparición de la barrera de energía libre.

En resumen, podemos decir que un equilibrio entre las interacciones locales y terciarias se traduce a su vez en un equilibrio entre energía y entropía. La contribución de ambas interacciones debe ser tal que ambos estados se diferencien lo suficientemente tanto en su energía como en su entropía, de manera que aparezca la barrera de energía libre característica de una transición de dos estados.

Estos resultados, por una parte, son de especial importancia a la hora de establecer la función matemática de un potencial basado en la topología del estado nativo, dada

la influencia directa que tiene en las propiedades termodinámicas de la transición que se describe. Esto debe tenerse en cuenta a la hora de utilizar funciones matemáticas derivadas en campos de fuerza de la mecánica molecular. En estos casos, a menudo se utilizan coeficientes armónicos elevados en la interacción de los términos de torsión. Esto puede generar cadenas muy rígidas que, como hemos visto, pueden llevar a conclusiones erróneas, como han observado también otros autores^{76,89}.

Por otra parte, hemos determinado que 2GB1 presenta una transición de dos estados en un amplio intervalo de contribuciones de energía terciaria. En el caso de que la intensidad relativa de las interacciones sea la misma para las de largo y corto alcance, esta proteína presenta un 50 % de cada una de ellas, lo cual está dentro del intervalo propuesto. Analizando una base de datos de 1590 proteínas, hemos visto que la mayoría presentan una proporción de contactos no locales entre un 50 y un 60 %. Esta característica general de la estructura de las proteínas podría explicar por qué la mayoría de las transiciones de plegamiento de proteínas con un único dominio han sido caracterizadas como transiciones de dos estados. Sin embargo, en proteínas pequeñas la proporción de contactos de largo alcance disminuye, lo cual podría ser indicativo de un comportamiento diferente de estas proteínas a la hora de adquirir la estructura del estado nativo.

Hemos tomado entonces un conjunto de proteínas pequeñas con distintas topologías y cuyas características estructurales corresponden, en algunos casos, a las de la mayoría (entre un 50 y un 60 % de contactos no locales) y en otros, a las de proteínas con menos de un 50 % de contactos terciarios. Para las proteínas que presentan entre un 50 y un 60 % de contactos de largo alcance, la transición que tiene lugar es una transición de dos estados, como indican los datos experimentales. Para las proteínas con una proporción de estructura terciaria muy por debajo del 50 %, como son 1L2Y y 1BBL, la transición resulta ser de tipo *downhill*. En el caso de proteínas con una contribución a la energía total por parte de las interacciones no locales cerca del 50 %, hemos observado tanto plegamientos sin barrera (1VII y 2PDD) como de dos estados (1W4E).

Estos resultados de nuestro modelo también coinciden con los resultados experimentales observados para cada una de estas proteínas.

Dado que la rigidez de la cadena puede modificar en algunos casos las propiedades de la transición, hemos realizado un análisis del efecto de la rigidez sobre la transición de todo el conjunto de proteínas estudiado. Comprobamos así que la coincidencia de los resultados de nuestras simulaciones con los datos experimentales no es producto de la parametrización del potencial. En el caso de las proteínas que presentan un plegamiento sin barrera, las características generales de la transición no se modifican al variar la rigidez dentro de intervalos razonables. Lo mismo ocurre en el caso de las proteínas con un plegamiento de dos estados, siempre y cuando la proporción entre las interacciones locales y de largo alcance permita el equilibrio necesario entre energía y entropía ya mencionado.

Los resultados obtenidos nos permiten llegar a varias conclusiones importantes. Dado que con un potencial basado en la topología del estado nativo hemos podido reproducir el perfil experimental de energía libre a T_m^* , se confirma el principio de frustración mínima, según el cual las interacciones que dirigen el plegamiento son las mismas que estabilizan el estado nativo. Una de las consecuencias que, según Bryngelson y colaboradores, tiene el principio de frustración mínima, es que la transición resulta ser de dos estados⁷. Sin embargo, utilizando el mismo potencial, también hemos descrito la transición de ciertas proteínas como un proceso sin barrera. Por lo tanto, la existencia de este tipo de proteínas no contradice el principio de frustración mínima. Por otra parte, hemos visto que el hecho de que la mayoría de las proteínas se plieguen de la misma manera probablemente se debe a que presentan propiedades estructurales semejantes y que la presencia de proteínas con un plegamiento sin barrera se puede explicar en términos de distintas características topológicas. Por último, queda claro que es la topología misma del estado nativo la que determina el tipo de transición que tiene lugar y no tanto la intensidad relativa de un tipo de interacciones con respecto de otro.

Sin embargo no es fácil establecer criterios topológicos unívocos que diferencien un tipo de proteína de otro. En general, podemos decir que aquellas proteínas con un porcentaje de energía de largo alcance en el estado nativo muy por debajo del 50 %, probablemente presenten un plegamiento sin barreras. Pero en proteínas con una proporción de estructura terciaria cercana al 50 % no puede hacerse esta predicción. Hemos visto que proteínas muy similares estructuralmente presentan comportamientos diferentes: 2PDD con un 42 % de energía no local y 1W4E con un 47 %. Además, hemos llegado a la conclusión de que, para que exista una barrera de energía libre entre el estado nativo y el desnaturalizado, no sólo es necesario cierto equilibrio entre las interacciones locales y las de largo alcance, sino también dentro del conjunto de interacciones de largo alcance. Las proteínas pequeñas que presentan una distribución de contactos de largo alcance con un marcado pico para los contactos helicoidales se pliegan según una transición tipo *downhill*. Por el contrario, las proteínas que presentan un plegamiento de dos estados tienen una distribución de contactos más homogénea a lo largo de todos los valores de órdenes de contacto. Queda claro que existe una relación entre la proporción de estructura terciaria respecto del total y el tipo de transición.

Podemos establecer entonces algunas características estructurales generales que diferencian un tipo de proteínas de otros. Sin embargo, el estudio realizado a partir de modificaciones en el mapa de contactos de 1W4E deja claro que no se puede establecer un límite absoluto entre un tipo de proteínas y otro. En el caso de proteínas donde las características estructurales están cerca pero no incluidos en la tendencia general, la distribución de los contactos uno a uno en el mapa de contactos hace que la transición sea de un tipo u otro. Pero sí podemos decir que la cooperatividad de una transición de plegamiento está relacionada con una distribución equitativa de los contactos a lo largo de la secuencia. Cuando un tipo de contactos predomina sobre otro la cooperatividad disminuye y puede tener lugar un plegamiento sin barrera. Queda clara así la aplicabilidad de los potenciales basados en la topología del estado nativo al estudio del

proceso de plegamiento. En el siguiente Capítulo veremos si también podemos utilizar estos potenciales para estudiar procesos de agregación.

Capítulo 5

Topología y agregación de proteínas.

En el capítulo anterior hemos estudiado cómo la topología del estado nativo de proteínas pequeñas determina en buena medida el tipo de transición de plegamiento que tiene lugar. Ya hemos comentado en la Capítulo 1 que las características de la transición de plegamiento parecen ser además de gran importancia para evitar el proceso de agregación. Como ya hemos explicado, este proceso de agregación se ve favorecido debido a que las condiciones celulares no siempre son las más apropiadas para el plegamiento de proteínas¹. La elevada concentración y la presencia de otras moléculas en el medio, que provocan un aumento aún mayor de la concentración efectiva de proteína, hacen que el plegamiento se vea dificultado. El proceso de agregación es un proceso indeseado que impide el plegamiento correcto de las proteínas y, por tanto, anula su función, provocando las ya mencionadas *enfermedades del plegamiento*. Algunas enfermedades degenerativas como el Parkinson o el Alzheimer están asociadas a la presencia de agregados con cierto orden estructural. La tendencia a la agregación parece ser una propiedad general de las proteínas, pero en los organismos vivos existen mecanismos que controlan el proceso de plegamiento de manera que esta tendencia está minimizada^{2,45}.

Aparte de la presencia de otras macromoléculas que regulan el proceso de plegamiento², se ha propuesto que la mayoría de las proteínas han sido seleccionadas para presentar un plegamiento de dos estados precisamente para evitar la agregación. Par-

tiendo de la hipótesis de que en todo proceso de agregación interviene la presencia de un intermedio de plegamiento, dado que en un proceso de dos estados la población de configuraciones con energía, y por tanto estructura, intermedia entre los estados nativo y desnaturalizado es mínima, la probabilidad de agregación también es mínima^{2,51}. Por otra parte, procesos de plegamiento sin barrera que tienen lugar de manera rápida también podrían haber sido seleccionados para evitar la agregación²¹, en este caso por motivos cinéticos. Hemos visto que con un potencial basado en la topología del estado nativo se pueden reproducir las características termodinámicas de la transición de plegamiento. Por lo tanto, utilizando el mismo potencial, podemos estudiar cómo afecta el tipo de transición que tiene lugar a la formación o no de agregados. Para ello, en la primera parte de este capítulo desarrollamos una ampliación del potencial de tipo $G\bar{o}$ utilizado hasta ahora para comprobar cuál es el papel que juega la presencia o no de una barrera de energía libre entre el estado plegado y el desnaturalizado en un posible proceso de agregación.

Por otro lado, ya hemos comentado que utilizar un potencial basado en la topología del estado nativo no es del todo adecuado para estudiar la agregación, debido a la ausencia de frustración. Como se ha explicado en el Capítulo 1, la agregación de proteínas parece que se debe, principalmente, a las interacciones de tipo genérico presentes en las proteínas. La agregación es, por lo tanto, producto de cierto grado de frustración en el sistema. Si bien cierto tipo de agregados debidos a un proceso de intercambio de dominios pueden estudiarse utilizando los potenciales basados en la topología del estado nativo¹¹², lo más adecuado es utilizar potenciales que incluyan interacciones de tipo genérico⁹⁷⁻¹⁰⁶. Por eso, en la segunda parte de este capítulo modificamos el potencial de tipo $G\bar{o}$ utilizado hasta el momento de dos maneras diferentes. Por un lado, introducimos un tipo de interacción local no basado en las distancias de corto alcance en el estado nativo, manteniendo el mapa de contactos de la estructura terciaria y el potencial hasta ahora utilizado para las interacciones de largo alcance. Esta modificación

proporciona una mayor flexibilidad a la cadena, que da lugar a un grado de frustración lo suficientemente elevado como para favorecer, en cierta medida, la formación de agregados. En segundo lugar, introducimos a modo de perturbación un potencial de campo medio para las interacciones no locales, manteniendo el potencial de tipo $G\ddot{o}$ para las interacciones locales. En este caso, la frustración del sistema se debe a las características de la secuencia de aminoácidos, de manera que el embudo de plegamiento definido por este tipo de potencial deja de ser ideal.

5.1. Potencial basado únicamente en la topología del estado nativo.

En esta sección hemos realizado simulaciones con el potencial descrito hasta ahora, pero extendidas a sistemas con varias cadenas de proteína. Como ya hemos dicho en el Capítulo 1, aunque no sea del todo realista, en parte es lógico pensar que las interacciones no locales que se producen entre dos moléculas son equivalentes a aquéllas que tienen lugar dentro de cada molécula. De hecho, un mecanismo frecuente de agregación es a través de un intercambio de dominios entre dos o más cadenas de la misma proteína. Por lo tanto, en esta parte de la Tesis hemos utilizado un potencial de tipo $G\ddot{o}$ “cruzado”, que explicamos a continuación, para estudiar cómo de determinante es la presencia de una barrera de energía libre a la hora de evitar la agregación de proteínas. Además, podemos comprobar si con este potencial se puede reproducir la estructura de algunos agregados detectados experimentalmente, producidos por un intercambio de dominios.

5.1.1. Potencial y condiciones de simulación utilizados.

Al pasar a un sistema de varias cadenas es necesario definir la interacción entre ellas. Como en los capítulos anteriores, en este caso hemos considerado también interacciones locales y de largo alcance. Las primeras se producen entre los aminoácidos

i e $i \pm 2$ y entre i e $i \pm 3$ de cada cadena. Las interacciones de largo alcance de cada cadena polipeptídica consigo misma, definidas por el mapa de contactos del estado nativo, tienen lugar entre los aminoácidos i y $j \geq i + 4$, y son las que llamamos interacciones intramoleculares o intra-cadena. Entre dos moléculas, las interacciones intermoleculares o inter-cadena se producen entre el aminoácido i de una cadena y los aminoácidos $j' \geq i' + 4$ de la otra, donde el símbolo $'$ indica la pertenencia a una segunda cadena. Para determinar entre qué pares de aminoácidos de dos cadenas diferentes existe interacción, se parte del mapa de contactos de la estructura nativa de una sola cadena, que es la base del potencial utilizado a lo largo de este trabajo. Si dos aminoácidos i y j forman un contacto en el estado nativo, por supuesto también lo hacen i' y j' y, además, definimos el mismo potencial de atracción entre i y j' y entre i' y j . Para un sistema de dos cadenas, por ejemplo, el mapa de contactos cruzado aparece entonces dividido en dos zonas, marcadas en la Figura 5.1. Por un lado, aparecen los contactos de cada cadena consigo misma, que son los del estado nativo de la proteína (cuadrados **a** de la Figura 5.1). Por otra parte, se definen los contactos inter-cadena, marcados en los cuadrados **b** de dicha figura. Vemos que en un potencial de tipo $G\bar{o}$ cruzado, los contactos intermoleculares definen un mapa igual al del estado nativo, aunque sin los contactos locales. En sistemas de más cadenas se obtendrían mapas equivalentes. El potencial de interacción es el mismo definido hasta ahora [ecuación (2.1), $a = 0.7 \text{ \AA}$] y el mapa de contactos se define a partir del empaquetamiento de las cadenas laterales de los aminoácidos con $d_{max} = 4.5 \text{ \AA}$. Para el cálculo de la energía según la ecuación (2.1) se utiliza la misma distancia d_{ij}^{nat} para los pares $i - j$, $i' - j$, $i - j'$ e $i' - j'$.

Una variable a tener en cuenta cuando se hacen simulaciones de sistemas de más de una cadena es la concentración. Para poder controlar este parámetro las simulaciones se llevan a cabo en lo que se llama una caja de simulación¹¹⁷. Dado que esto puede provocar efectos superficiales no deseados, se utilizan condiciones periódicas de contorno mediante el criterio de mínima imagen explicado en el Capítulo 2. Las dimen-

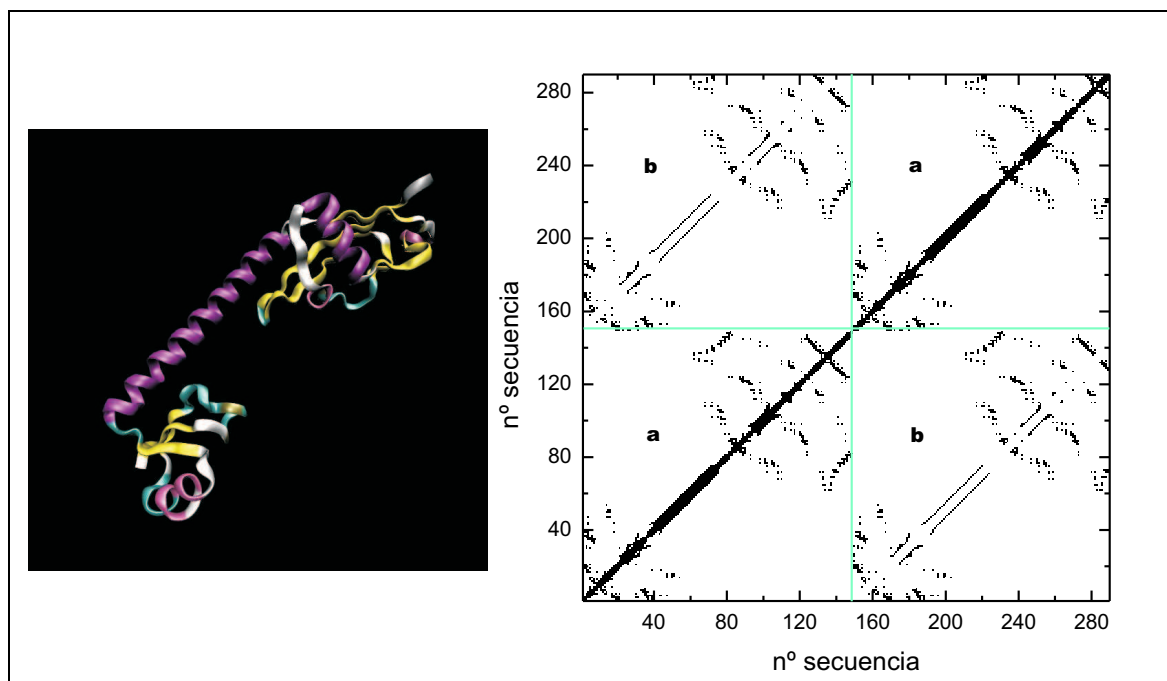


Figura 5.1: Mapa de contactos cruzado para un sistema de dos cadenas de la proteína representada (figura realizada con VMD⁶⁹). Las líneas azules indican el final de una cadena y el comienzo de la siguiente. La numeración de la secuencia es continuada, de manera que no se comienza en 1 otra vez en la segunda cadena. Los cuadrados **a** corresponden a las interacciones intra-cadena y los **b** a las inter-cadena.

siones de esta caja se eligen de manera que la concentración sea de 100 mg/mL¹⁰⁸. Esta concentración es bastante superior a la concentración que se utiliza en la mayoría de los experimentos, pero menor que la que presentan las proteínas en la célula, que está en torno a 200-300 mg/mL^{150,151}. Utilizando una concentración elevada favorecemos el proceso de agregación, que es el que queremos estudiar.

Hemos realizado simulaciones de sistemas de 2 y 4 cadenas para diferentes proteínas. Por un lado, hemos utilizado la proteína con código PDB 2GB1, que es una proteína que hemos estudiado en profundidad en capítulos anteriores de este trabajo y que presenta un plegamiento de dos estados. La hemos incluido en este estudio para comprobar si este tipo de plegamiento tiene alguna implicación en el proceso de agregación. Con el mismo objetivo, hemos utilizado también el inhibidor 2 de la quimotripsina. Esta proteína presenta también un plegamiento de dos estados y existen sobre ella estudios de

agregación experimentales^{48,152} validados con modelos teóricos sencillos¹⁵³. En este caso hemos utilizado la estructura con código PDB 1YPB (que difiere ligeramente en secuencia de la recogida bajo el código 2CI2, que es la que hemos usado en el capítulo anterior), ya que es la proteína para la cual se han obtenido los mencionados datos experimentales y podemos así establecer una relación más directa entre éstos y nuestros resultados. Por otra parte, hemos estudiado también sistemas de 2 y 4 cadenas de la proteína con código PDB 1BBL. Según nuestro modelo, y parece ser que también experimentalmente, esta proteína pliega siguiendo un proceso sin barrera. La hemos incluido en el conjunto de proteínas estudiadas para determinar cómo afecta al proceso de agregación la ausencia de barrera de energía libre entre los estados plegado y desnaturalizado. Con el objetivo de comprobar si los resultados obtenidos con 1BBL son producto del tipo de transición o efecto de la estructura de la proteína en la cual se basa el potencial, hemos incluido en el estudio el mutante 1W4E, que es un homólogo estructural de 1BBL pero que presenta una transición de primer orden cuando se pliega, como hemos comprobado también con nuestro modelo en el Capítulo 4. Los mapas de contactos para sistemas de dos cadenas de cada una de estas proteínas son los de la Figura 5.2.

5.1.2. Sistemas de dos y cuatro cadenas.

Con las condiciones en la metodología de simulación explicadas en el apartado anterior hemos llevado a cabo varias simulaciones de sistemas de 2 y 4 cadenas de las cuatro proteínas presentadas anteriormente, según lo explicado en el Capítulo 2. Para tener una idea inicial de las propiedades de la transición de estos sistemas hemos calculado, a partir de los resultados de dichas simulaciones, de capacidad calorífica frente a la temperatura. En la Figura 5.3 vemos los resultados obtenidos comparados con los resultados de sistemas de una cadena. Para las proteínas 1W4E, 2GB1 y 1YPB, es decir, aquéllas que presentan un plegamiento de dos estados, las curvas de capacidad calorífica no presentan modificaciones importantes al aumentar el número de cadenas. Para estas

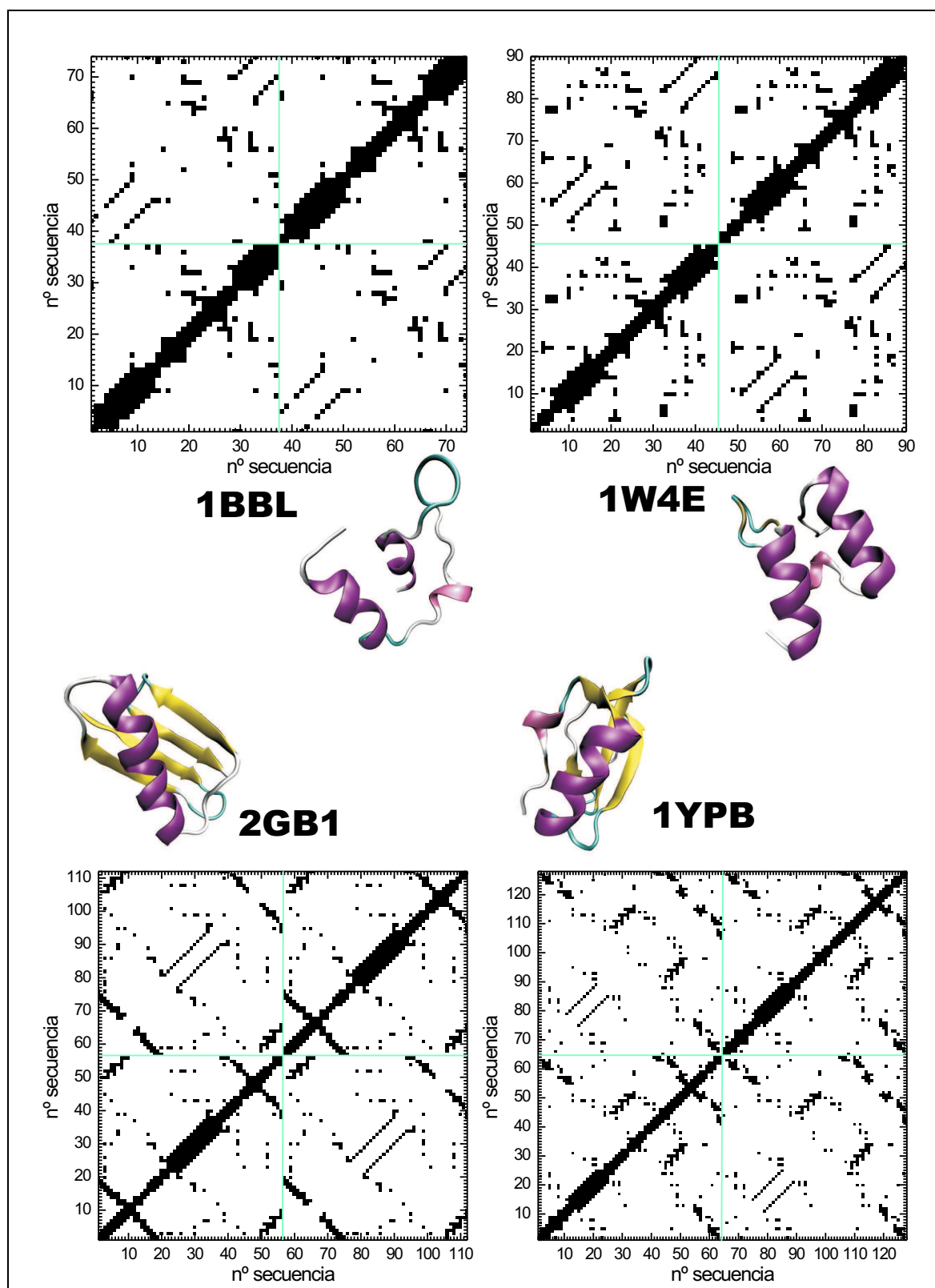


Figura 5.2: Representación del mapa de contactos cruzado para sistemas de dos cadenas de las proteínas indicadas (representación realizada con VMD⁶⁹). Las líneas azules indican el final de una cadena y el comienzo de la siguiente.

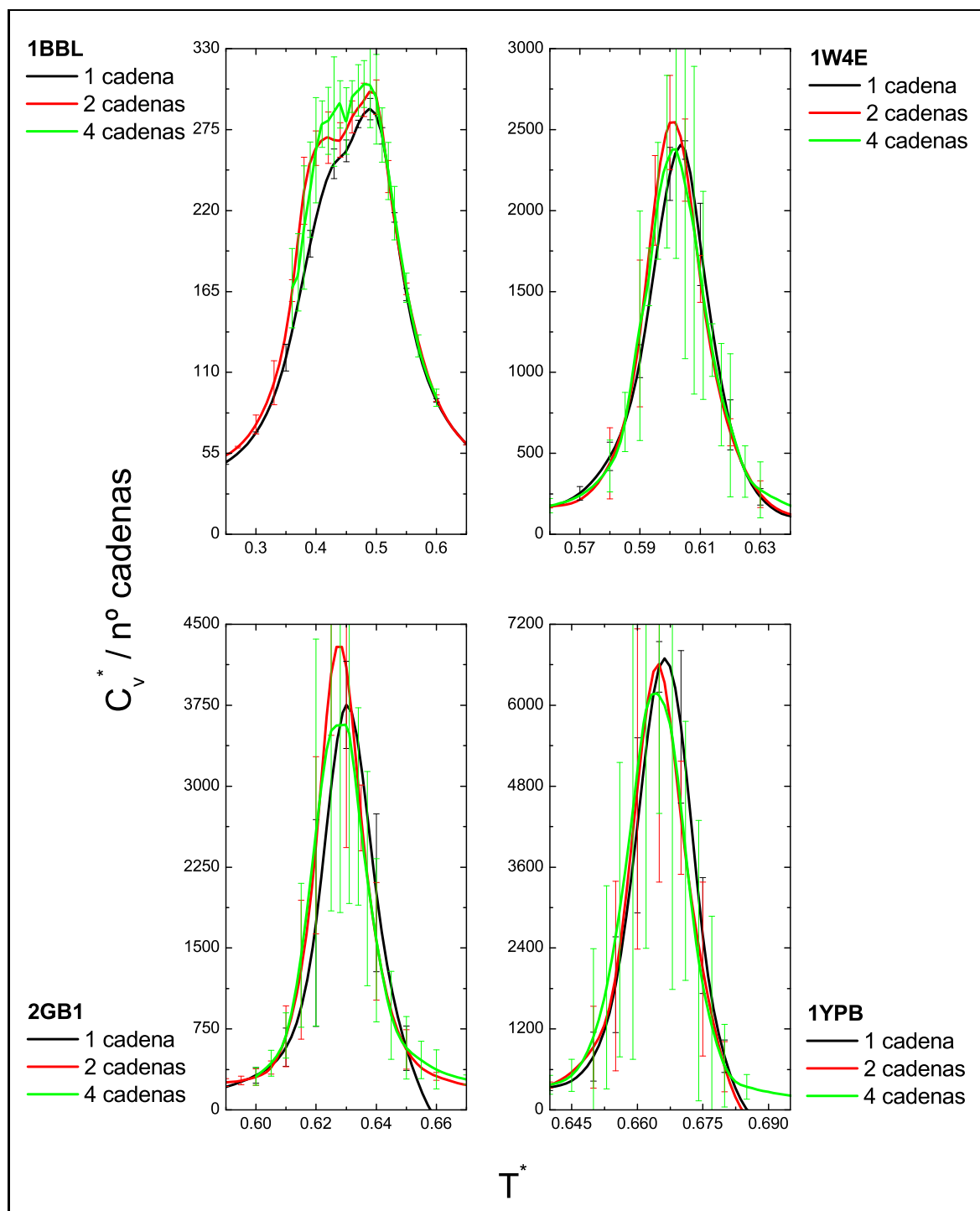


Figura 5.3: Capacidad calorífica para sistemas de una, dos y cuatro cadenas de las proteínas indicadas.

Tabla 5.1: Valores de la temperatura de transición T_m^* , deducida a partir del máximo de capacidad calorífica, para sistemas de una, dos y cuatro cadenas. El error es de la unidad en la última cifra.

| | 1 cadena | 2 cadenas | 4 cadenas |
|------|----------|-----------|-----------|
| 1BBL | 0.49 | 0.50 | 0.48 |
| 1W4E | 0.605 | 0.60 | 0.602 |
| 2GB1 | 0.63 | 0.63 | 0.628 |
| 1YPB | 0.665 | 0.665 | 0.665 |

tres proteínas, las curvas de capacidad calorífica cambian ligeramente de forma y altura al cambiar el número de cadenas. Sin embargo, las propiedades generales (especialmente la presencia de un solo pico estrecho) y los valores de la temperatura de transición (cuyos valores aparecen en la Tabla 5.1) se mantienen al aumentar el número de cadenas. Por el contrario, en el caso de 1BBL, las propiedades de la transición varían a medida que aumenta el número de cadenas, apareciendo de forma diferenciada en sistemas de varias cadenas un segundo pico en C_v^* a menor temperatura que T_m^* de la cadena aislada.

Como venimos observando a lo largo de esta Tesis, para caracterizar la transición desde un punto de vista termodinámico es necesario calcular los perfiles de energía libre a T_m^* . A partir de ellos podemos determinar si las características termodinámicas de la transición, en cuanto a la presencia o ausencia de una barrera de energía libre, se mantienen cuando aumenta el número de cadenas. En la Figura 5.4 representamos estos perfiles de energía libre. Para los sistemas de cuatro cadenas, encontramos un problema numérico derivado de la degeneración inmensamente elevada de los estados de alta energía, que impide el cálculo específico de la energía libre para estas energías. Pero podemos extrapolar los resultados y deducir el resto del perfil de energía libre por comparación con lo que sucede en otros sistemas. En la Figura 5.4 hemos marcado la extrapolación realizada con línea discontinua.

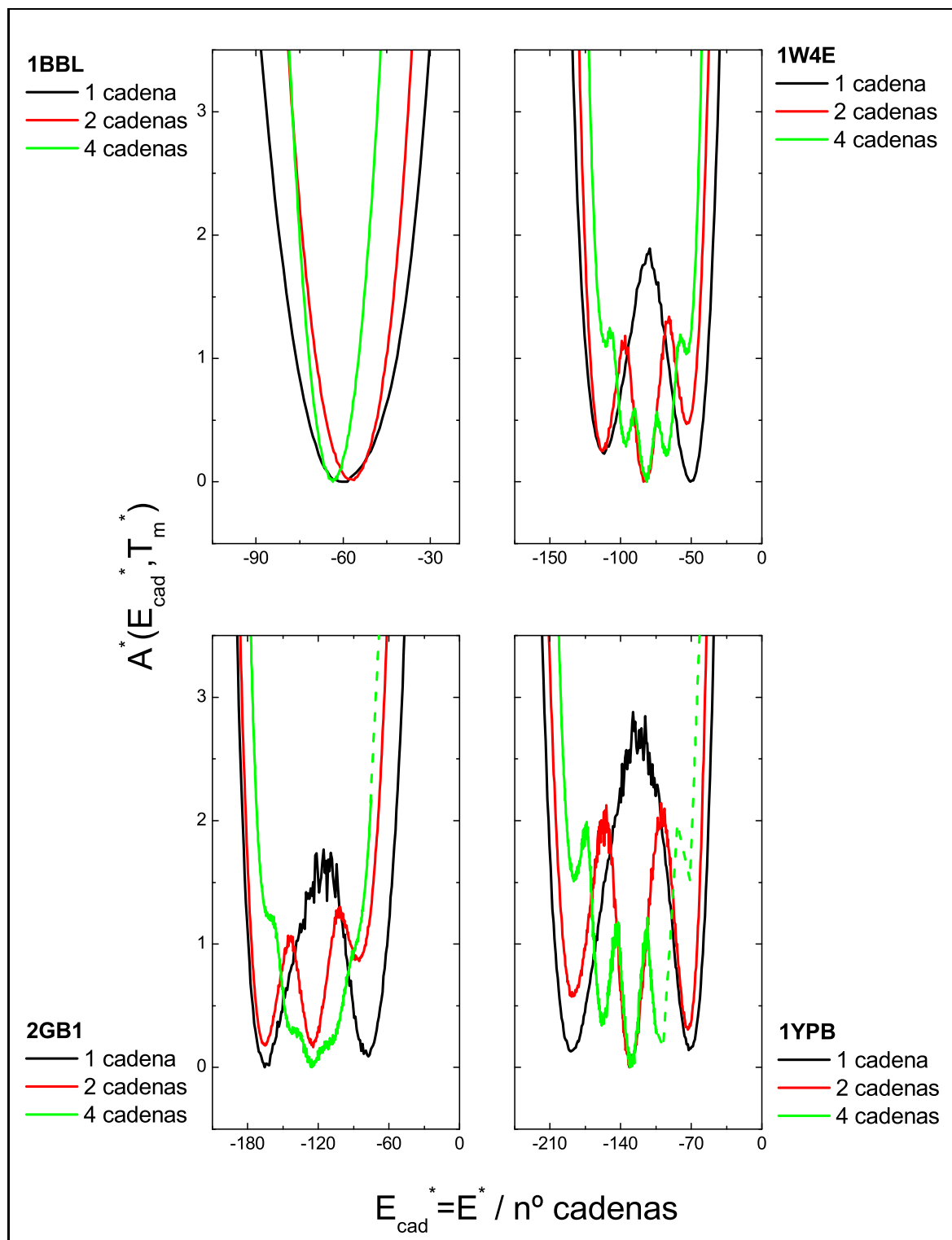


Figura 5.4: Perfiles de energía libre a T_m^* para sistemas de una, dos y cuatro cadenas. La línea discontinua corresponde a una extrapolación de los resultados.

Aunque inicialmente pueda parecer que la transición ha sufrido modificaciones importantes, como explicamos a continuación, en general las propiedades de la transición no varían. Para aquellas proteínas que presentan una transición de primer orden, observamos un aumento en el número de mínimos en el perfil de energía libre cuando aumenta el número de cadenas. Estos mínimos no corresponden a intermedios de plegamiento, sino al equilibrio adicional entre unas cadenas ya plegadas y otras que permanecen en el estado desnaturalizado. El plegamiento de todas las cadenas está en equilibrio a la temperatura de transición, dando lugar a todos los estados posibles: 1 cadena plegada y las demás desnaturalizadas, 2 cadenas plegadas, ... En el caso de 1YPB y 1W4E, entre todos los estados presentes en los sistemas de dos y cuatro cadenas aparece una barrera de energía libre notable. Sin embargo, para sistemas de cuatro cadenas de 2GB1, la barrera entre los diferentes estados desaparece, aunque podemos distinguir varios hombros en el perfil de energía libre, debidos al solapamiento de los distintos estados. Por otra parte, la transición de 1BBL tampoco se modifica con el aumento de concentración: se trata de una transición sin barrera en los tres casos.

Dado que, en general, no observamos ninguna modificación importante en las propiedades termodinámicas del tránsito, es de esperar que el estado de mínima energía al que se llega sea el estado nativo para cada una de las cadenas. Con el objetivo de confirmar esta hipótesis hemos realizado un análisis del valor medio de *rmsd* con respecto al estado nativo de las conformaciones presentes a una temperatura T_{str}^* menor que T_m^* , equivalente en todos los casos. Si el valor medio de esta propiedad difiere mucho para una misma proteína al modificar el número de cadenas, es posible que se deba a la presencia de agregados. Excepto en los casos de cuatro cadenas de 1W4E, 2GB1 y 1YPB, hemos utilizado un valor de temperatura $T_{str}^* = 0.75T_m^*$. Esta temperatura es lo suficientemente baja como para asegurarnos de que hemos encontrado la configuración más estable. En las simulaciones de 4 cadenas de 1W4E, 2GB1 y 1YPB no se ha llegado hasta temperaturas tan bajas (ver Figura 5.3), por lo que tomamos la menor tempe-

Tabla 5.2: Valor medio de *rmsd* respecto a la estructura nativa de las configuraciones recogidas a una temperatura $T_{str}^* = 0.75T_m^*$ para los sistemas de una, dos y cuatro cadenas (ver texto).

| | n° cadenas | | |
|------|---------------|---------------|---------------|
| | 1 | 2 | 4 |
| 1BBL | 2 ± 1 | 2 ± 1 | 3 ± 2 |
| 1W4E | 1.7 ± 0.6 | 1.6 ± 0.6 | 2.2 ± 0.7 |
| 2GB1 | 0.6 ± 0.1 | 0.7 ± 0.2 | 1.2 ± 0.8 |
| 1YPB | 0.8 ± 0.3 | 0.8 ± 0.5 | 1.1 ± 0.4 |

ratura a la que se ha realizado una simulación en cada una de ellas, que se encuentra en el intervalo $(0.90T_m^*; 0.95T_m^*)$. Los resultados, presentados en la Tabla 5.2, indican que, para los sistemas de varias cadenas, el valor medio de *rmsd* es bastante parecido al que se alcanza con una sola cadena. De esta forma sabemos que en todos los casos las proteínas adquieren el estado nativo a baja temperatura. Dado que la temperatura utilizada en el caso de cuatro cadenas es mayor que en los demás en comparación con T_m^* , es razonable encontrar valores de *rmsd* ligeramente superiores. Sin embargo, también en estos casos el valor medio de *rmsd* es lo suficientemente bajo como para poder afirmar que se ha adquirido la estructura del estado nativo, lo que confirmamos con la inspección visual de las conformaciones registradas a esta temperatura a partir de las simulaciones realizadas.

Los resultados obtenidos responden claramente a las características del potencial. Dado que se trata de un sistema no frustrado, es inevitable la adquisición del estado nativo. Cuando la simetría de estas proteínas lo permite, se producen interacciones intermoleculares adicionales entre las cadenas plegadas. Sin embargo, consideramos que esto no es un proceso de agregación propiamente dicho, sino simplemente una consecuencia del proceso de minimización de energía a bajas temperaturas. Las interacciones definidas dan lugar a una tendencia a maximizar el número de contactos, pero no a costa

de perder la estructura nativa.

Queda claro entonces que el proceso favorable en estos casos es siempre el plegamiento. Sin embargo, esto no excluye la posibilidad de que, a lo largo de la transición, exista cierta competencia entre la agregación y el plegamiento. De hecho, ya hemos comentado en el Capítulo 1 de esta Tesis que experimentalmente se ha comprobado que en la transición de plegamiento se pueden formar agregados a partir de conformaciones desnaturalizadas de la proteína en los que ésta queda atrapada pero que, posteriormente, se forma el estado nativo⁴⁸. Esta observación ha sido realizada, concretamente, a partir de estudios sobre el plegamiento de 1YPB.

Ni las curvas de capacidad calorífica ni los perfiles de energía libre nos dan información acerca de la competencia que pueda haber entre las interacciones inter e intramoleculares. Por lo tanto, hemos utilizado las ecuaciones (2.18), (2.19) y (2.20) del método WHAM para calcular el mapa de energía libre en función de dos propiedades que sí nos den idea del grado de interacción que tiene lugar entre cadenas frente al que tiene lugar dentro de cada cadena. Las propiedades que hemos utilizado son el número de contactos intramoleculares por cadena y el número de contactos intermoleculares formados.

1BBL es la única de las proteínas que estamos estudiando que presenta una transición de plegamiento sin barrera. Además, hemos visto que su curva de capacidad calorífica es la única de las cuatro que se modifica apreciablemente con el aumento del número de cadenas, apareciendo un segundo pico a menor temperatura. Si analizamos los mapas de energía libre de las transiciones de los sistemas de 2 y 4 cadenas de esta proteína, podemos comprobar qué es lo que sucede a cada una de las temperaturas a las que C_v^* presenta un máximo. En la Figura 5.5 mostramos los mapas de energía libre a estas temperaturas en función de los contactos inter- e intra-cadena que, como hemos dicho, nos permiten estudiar la competencia entre la agregación y el plegamiento. A la temperatura de transición T_m^* , el mínimo de energía libre para los sistemas de dos

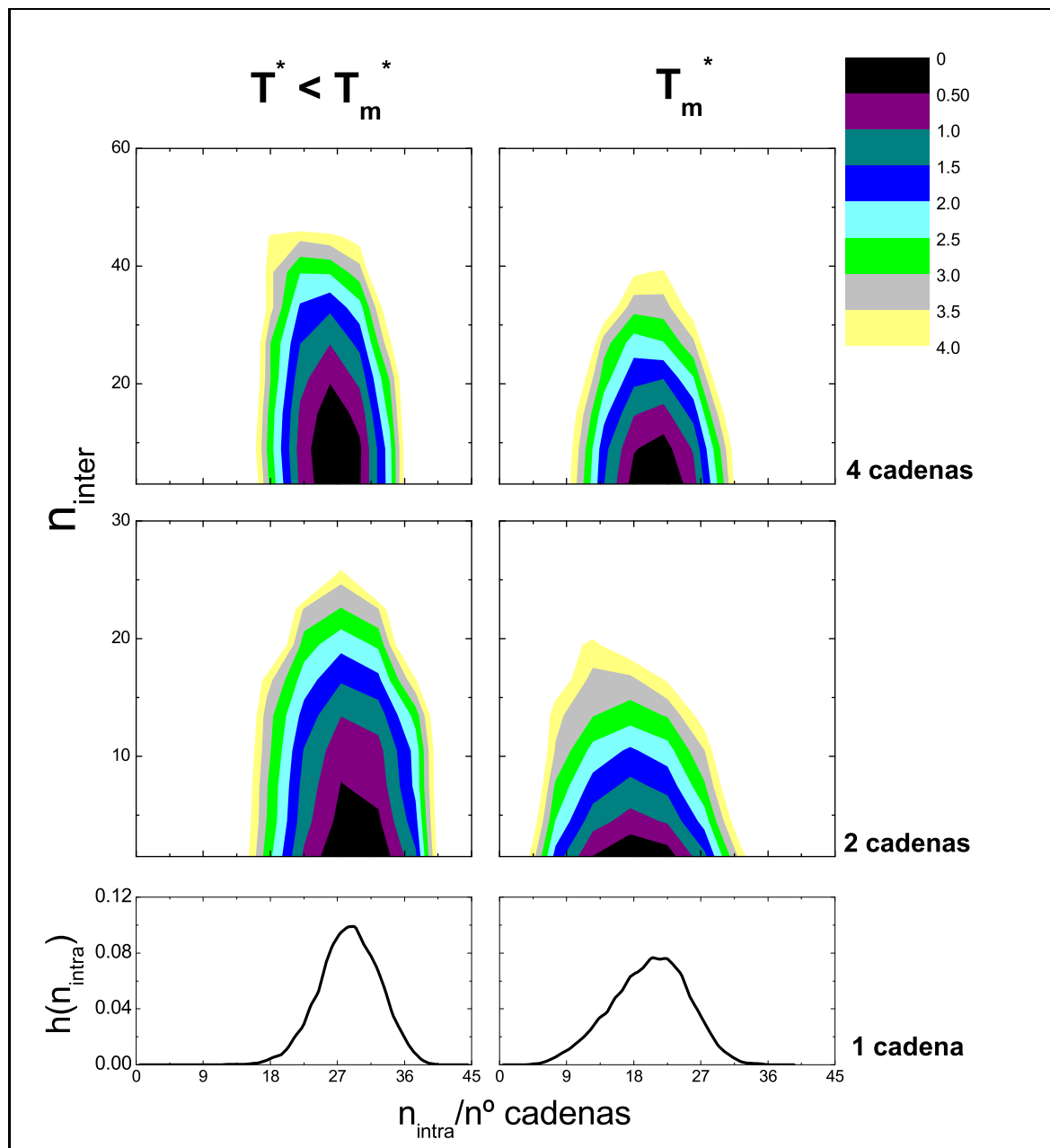


Figura 5.5: Mapas de energía libre para los sistemas de 2 y 4 cadenas de 1BBL en función de los contactos inter e intramoleculares formados a las temperaturas a las que la capacidad calorífica presenta un máximo. En la parte inferior se representan los histogramas de los contactos terciarios intramoleculares que están formados a esas temperaturas en los sistemas de una cadena.

y cuatro cadenas corresponde a conformaciones en que el grado de estructura terciaria que se adquiere es el mismo que en el caso de sistemas de una cadena. Es decir, esta transición corresponde al plegamiento gradual de la proteína. A menor temperatura, también se adquiere un grado de estructura terciaria equivalente al del sistema de una cadena. Pero además, a esta temperatura, el mínimo de energía libre se extiende hacia valores mayores de contactos inter-cadena que a la temperatura T_m^* . Esto quiere decir que a temperatura menor que T_m^* se produce un cierto proceso de asociación intermolecular. Según los datos de la Tabla 5.2, a bajas temperaturas las cadenas están en el estado nativo. En la transición a $T^* < T_m^*$ lo que sucede es que, a medida que la proteína adquiere el estado nativo, se producen también interacciones entre las cadenas, de manera que a baja temperatura lo que se obtienen son cadenas en el estado nativo adheridas unas a otras. En el caso de sistemas de cuatro cadenas de 1BBL, vemos que el número medio de contactos intramoleculares formados es ligeramente inferior que para los casos de una y dos cadenas. Sin embargo, la tendencia en este caso también es a adquirir el estado nativo y la asociación de las cadenas en este estado. Como hemos dicho, no consideramos que la formación de contactos entre cadenas plegadas corresponda, estrictamente, a un proceso de agregación. Sin embargo, es importante destacar que este proceso de asociación, en el caso de 1BBL, da lugar a un segundo pico en la capacidad calorífica, mientras que en los otros casos (1W4E, 2GB1 y 1YPB) no sucede así. Aunque el plegamiento es el proceso más favorecido, la ausencia de barrera en la transición de 1BBL hace que la asociación de cadenas suceda de manera simultánea a la adquisición de estructura terciaria.

Por otra parte, en el caso de las proteínas que sí presentan una barrera de energía libre entre los estados plegado y desnaturalizado, la transición no cambia sustancialmente. Las curvas de capacidad calorífica son equivalentes (Figura 5.3) en todos los casos y hemos visto que los perfiles de energía libre siguen siendo, mayoritariamente, los característicos de una transición de primer orden (Figura 5.4). Además, el valor medio

de *rmsd* a temperatura significativamente menor que T_m^* (Tabla 5.2) deja claro que el estado más estable es el nativo y que el efecto de la concentración no es suficiente para dar lugar a la agregación. Pero para poder saber si existe competición entre agregación y plegamiento a lo largo de la transición recurrimos de nuevo a los mapas de energía libre, representados en la Figura 5.6. Para los sistemas de dos cadenas vemos que, en todos los casos, aparecen varios mínimos de energía libre. Los tres que aparecen a valores bajos de número de contactos inter-cadena corresponden al estado completamente desnaturado, el plegamiento de una cadena y el plegamiento de las dos, respectivamente, al igual que en los perfiles de la Figura 5.4. Además, aparece otro mínimo de energía libre adicional (dos en el caso de 1YPB) que corresponde a un estado en el que hay un elevado número de contactos intermoleculares.

A partir de la visualización de las conformaciones registradas a esta temperatura podemos ver cuáles son las conformaciones que dan lugar a estos mínimos locales de energía libre. En el mapa de 1W4E aparece un estado formado por un conjunto de conformaciones cuyo número de contactos intramoleculares indica que ninguna de las dos cadenas está plegada. Este estado está formado por cadenas abiertas con muy poca estructura terciaria entre las que existe cierto número de interacciones inter-cadena. En el caso de 2GB1 y 1YPB, sin embargo, las conformaciones correspondientes a los mínimos locales de energía libre que aparecen son agregados producidos por un intercambio de dominios en que cada cadena individual presenta cierto grado de estructura terciaria. Es importante destacar aquí la escala de contactos inter-cadena, que es muy diferente para 1W4E con respecto a 2GB1 o 1YPB. Vemos entonces que estos mínimos locales correspondientes a estados agregados presentan muchos más contactos intermoleculares en 2GB1 y 1YPB que en 1W4E. Para 2GB1, el número de contactos intramoleculares por cadena en este agregado es equivalente al del estado en que una proteína está plegada y otra no, pero en este caso, cada cadena individual presenta un número menor de contactos intra-cadena que en el estado plegado. El conjunto de configuraciones es

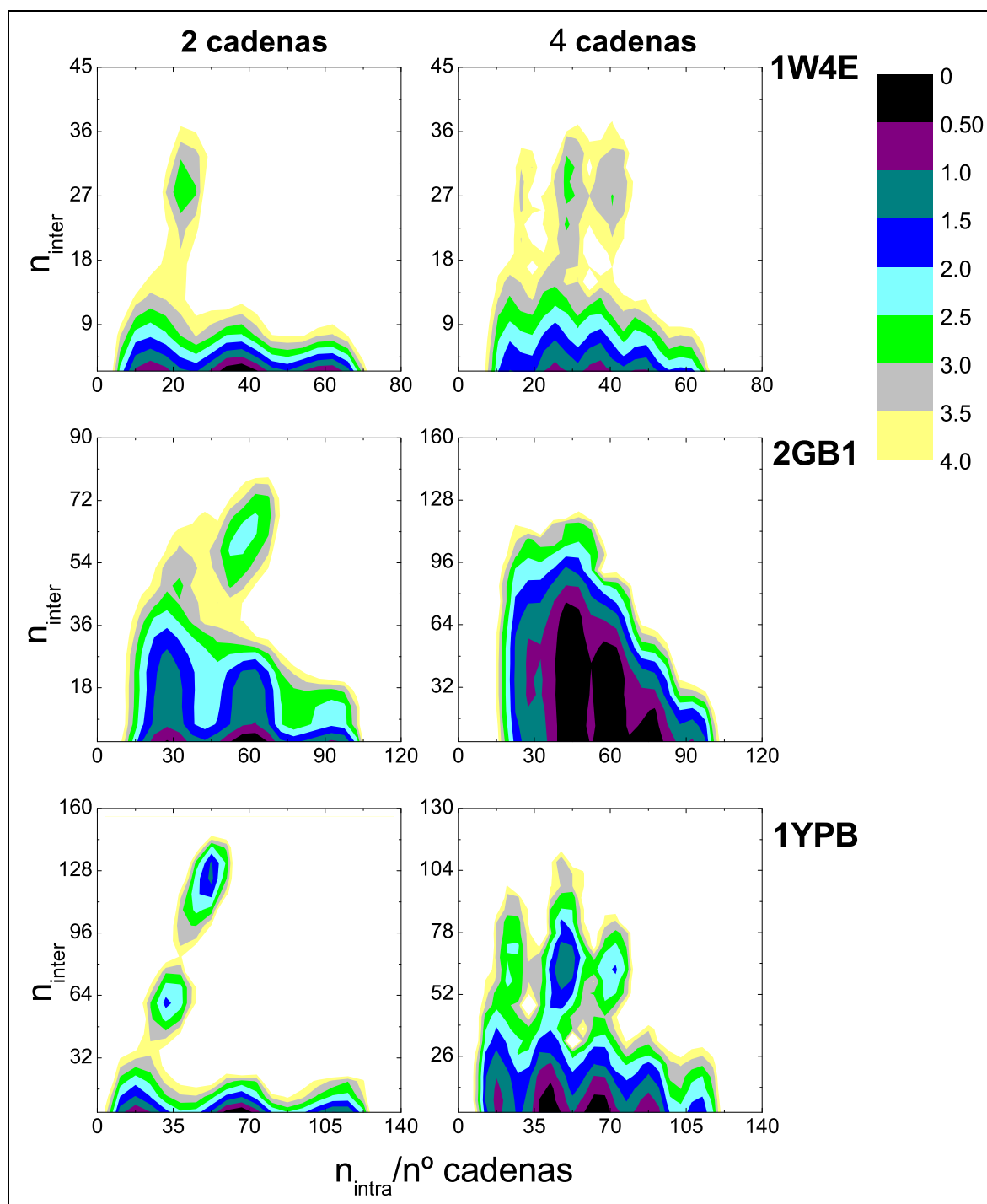


Figura 5.6: Mapas de energía libre para los sistemas de 2 y 4 cadenas de las proteínas indicadas en función de los contactos inter e intramoleculares formados a la temperatura de transición T_m^* .

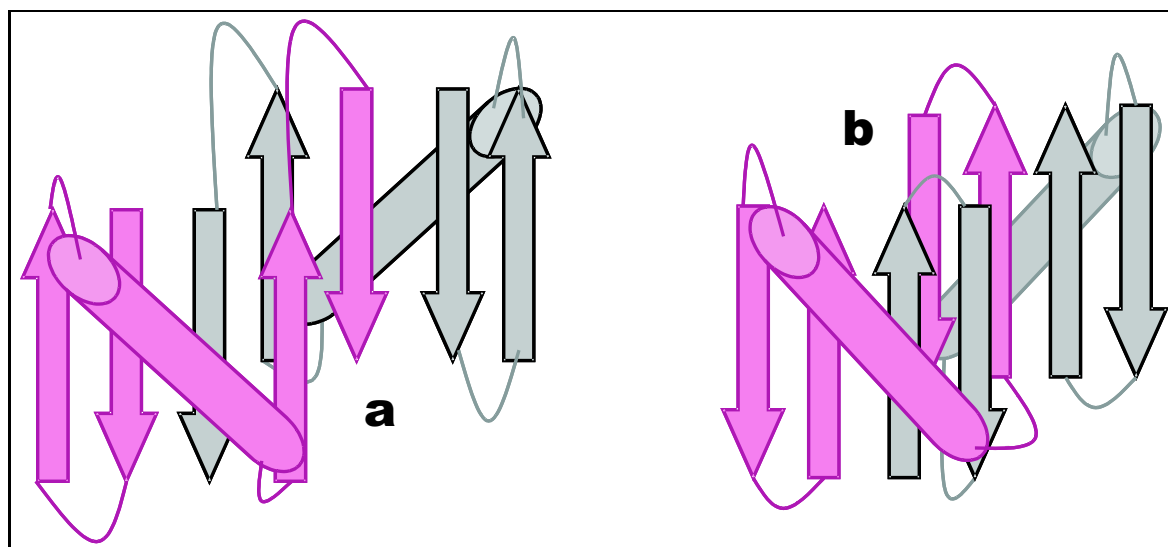


Figura 5.7: Representación esquemática de las conformaciones que dan lugar a un estado agregado por intercambio de dominios en 2GB1.

variado, dependiendo del grado de estructura terciaria que cada cadena adquiere. Para proporcionar una idea del tipo de estructuras que presentan la energía libre del mínimo local, en la Figura 5.7 representamos esquemáticamente algunas conformaciones representativas de este mínimo local que hemos observado en nuestras simulaciones. El intercambio de dominios se produce de dos maneras diferentes: por intercambio de una hebra, de manera que además ésta mantiene los contactos con su vecina dentro de la cadena (Figura 5.7a); o por intercambio de toda una horquilla, maximizando también el número de contactos entre cadenas (Figura 5.7b). El grado de estructura secundaria adquirido en cada cadena varía en todo el conjunto de configuraciones y aparecen algunas en que la hélice no está bien formada o en que una de las cadenas sólo ha adquirido la estructura nativa en el dominio que intercambia. Esto hace que el mínimo local sea ancho (ver Figura 5.6), englobando un gran número de configuraciones. Es interesante mencionar que, a partir de estudios experimentales sobre mutaciones de esta proteína, se ha identificado un mutante que da lugar a un dímero formado por intercambio de dominios. Este dímero se forma por el intercambio de la segunda horquilla β^{154} . Aunque las conformaciones que se detectan en nuestras simulaciones no son del todo iguales a

la estructura del dímero detectada experimentalmente, sí reflejan la relación entre la estructura y la tendencia a formar este tipo de dímeros.

En el caso de las simulaciones de dos cadenas para 1YPB, dependiendo del grado de estructura adquirido por cada cadena aparecen dos estados agregados diferentes, es decir, los dos mínimos locales de energía libre que observamos en la Figura 5.6. El mínimo que corresponde a un mayor número de contactos inter-cadena tiene también más contactos intramoleculares por cadena, indicando un mayor grado de estructura terciaria adquirido. Este mínimo corresponde a un conjunto de configuraciones en el que el intercambio de dominios que se produce es elevado. Este intercambio se produce en menor medida en las configuraciones que dan lugar al otro mínimo local de energía libre con un número de contactos intermoleculares menor. Para hacernos una idea visual de cómo se produce el intercambio, en la Figura 5.8a podemos ver un ejemplo de una configuración que presenta un número de contactos inter-cadena más elevado, mientras que el de menor número de contactos intermoleculares se muestra en la Figura 5.8b (la estructura nativa, referencia para la comparación, se muestra en la Figura 5.2). En el segundo caso, se produce el intercambio en menor medida, de manera que la interacción entre dos cadenas es equivalente a la formación del estado nativo en una cadena. Por el contrario, la disposición de las cadenas de la Figura 5.8a es equivalente a la formación del estado nativo en las dos cadenas. De esta forma aumenta tanto el número de contactos formados entre cadenas como el de contactos dentro de cada una de ellas, que presentan un grado de estructura nativa mayor. Estas proteínas dan lugar a este tipo de agregados porque la topología del estado nativo y la simetría estructural del mismo lo permite. En el caso de 1W4E la estructura del estado nativo hace que este mecanismo de agregación esté más impedido y que, por lo tanto, no lo observemos. Una diferencia importante entre la estructura de 1W4E con respecto de 2GB1 y 1YPB es que éstas últimas presentan una proporción elevada de hebras β en su estructura, mientras que 1W4E carece de estos elementos de estructura secundaria. A partir de nuestros resultados parece pues

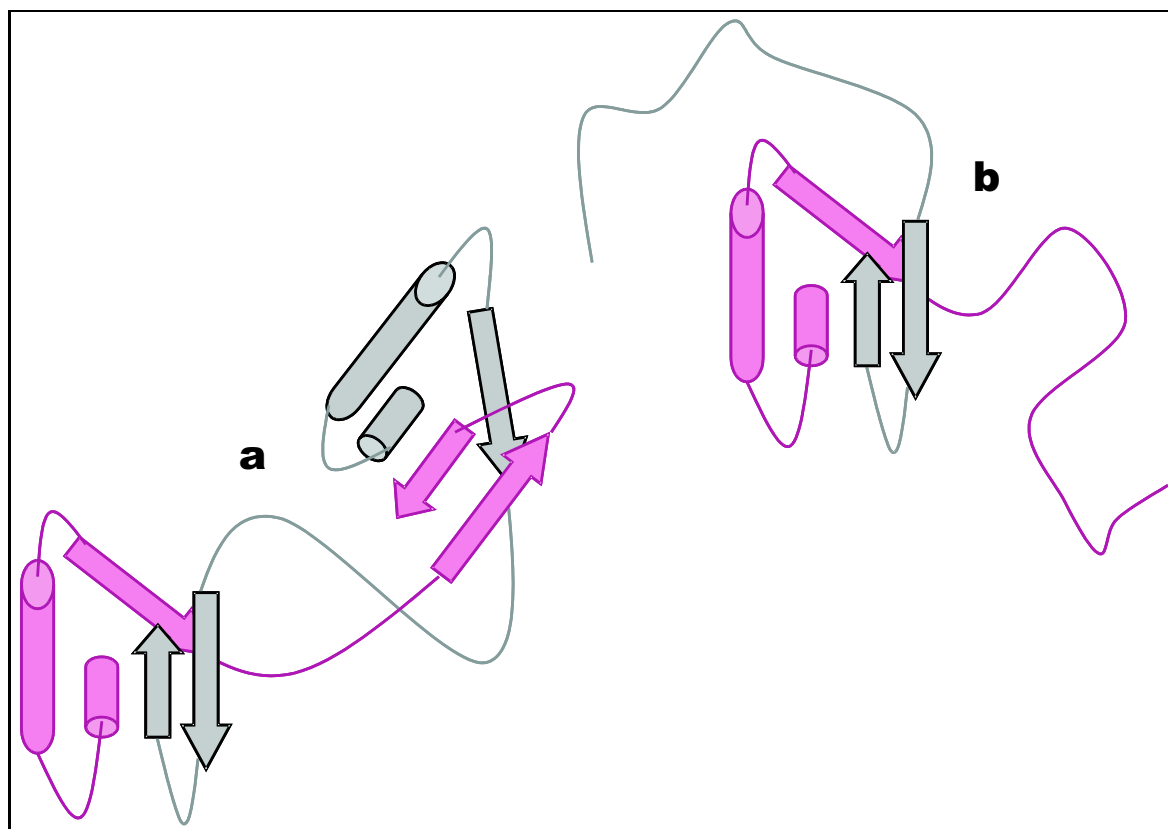


Figura 5.8: Representación esquemática de las conformaciones que dan lugar a estados agregados por intercambio de dominios en 1YPB. En **a** se representa el tipo de configuración que da lugar a un mayor número de contactos inter- e intra-cadena y en **b** el que da lugar a un número menor de ambos tipos de contactos.

razonable aventurar que la presencia de hebras β puede dar lugar a una mayor tendencia a formar dímeros por intercambio de dominios, lo cual podría estar relacionado con la formación de agregados β -amiloides en que la agregación se produce por interacción entre este tipo de elementos de estructura secundaria^{2,43}.

En los mapas de energía libre de los sistemas de cuatro cadenas aparecen también los mínimos correspondientes al plegamiento de las cadenas individuales (ver Figura 5.6). Además, también observamos la aparición de los agregados que hemos detectado en los sistemas de dos cadenas. Para 1W4E y 1YPB se forman agregados de dos cadenas, pudiendo estar las otras dos en el estado plegado o no. Los mínimos locales de energía libre se diferencian de la misma manera que en los sistemas de dos cadenas. Sin

embargo, estos mínimos no se diferencian tan claramente en el caso de 2GB1. En este caso, los mínimos de energía libre correspondientes a una, dos o tres cadenas plegadas, con el resto en el estado desnaturalizado, están solapando entre sí y con los mínimos que corresponden a configuraciones en que se han formado agregados por intercambio de dominios. El intercambio de dominios se puede producir entre dos o más cadenas, dando lugar una variabilidad mucho mayor de los estados agregados, que es lo que produce el solapamiento de los mínimos de energía libre y la desaparición de la barrera observada en la Figura 5.4. Esto se debe a la enorme simetría que presenta 2GB1, que permite que se produzca el intercambio de la horquilla $\beta 1$ de una cadena con otra a la vez que se intercambia la horquilla $\beta 2$ con una tercera cadena. De hecho, se han detectado experimentalmente estructuras tetraméricas de mutantes de esta proteína¹⁵⁵. Las conformaciones que aparecen en nuestra simulación no presentan parecido estructural con la estructura del tetrámero determinado experimentalmente. Sin embargo, que la proteína 2GB1, con este potencial de interacción, de lugar a tetrámeros, cosa que no ocurre con las otras proteínas, indica que es su estructura bastante simétrica lo que proporciona a esta proteína su versatilidad.

Otra observación importante que puede hacerse en los mapas de la Figura 5.6 es que los mínimos locales de energía libre con elevado número de contactos inter-cadena corresponden a estados en que la estructura terciaria de cada cadena no es la del estado nativo, es decir, los agregados se forman a partir del estado desnaturalizado. Esta afirmación viene reforzada por el hecho de que, entre el mínimo de energía libre del estado desnaturalizado y estos mínimos locales la barrera de energía libre es menor que entre éstos y los mínimos de las estructuras plegadas que aparecen en dichos mapas. Este resultado de nuestras simulaciones está de acuerdo con las observaciones experimentales de Silow y colaboradores⁴⁸. Esto permite establecer una relación entre el principio de frustración mínima y la agregación de proteínas, aunque el proceso más favorecido sea el plegamiento, como cabe esperar con un potencial de tipo Gō. Para que tenga lugar la

agregación en una proporción mayor es necesario que exista mayor competencia entre las interacciones que intervienen. Esto se consigue modificando el esquema de interacciones de manera que éste sea más genérico. En las secciones siguientes hemos llevado a cabo esta modificación de dos maneras diferentes.

5.2. Modificación de las interacciones locales.

En la sección anterior hemos utilizado el potencial basado en la topología del estado nativo estudiado a lo largo de esta Tesis, que es específico para cada estructura, para estudiar la posible competencia entre agregación y plegamiento. Aunque con este potencial se pueden reproducir resultados experimentales del plegamiento de proteínas, hemos visto que da información limitada en lo que al proceso de agregación se refiere. Es cierto que permite estudiar el papel que juega el tipo de transición en la agregación, pero, dadas sus características, describe sistemas demasiado poco frustrados, de manera que no observamos la formación de agregados a baja temperatura. En esta sección y en la siguiente hemos modificado parte del potencial de interacción de manera que las interacciones dejen de ser específicas de la estructura y pasen a ser generales para todas las proteínas. Se introduce así un grado de frustración tal que la agregación puede tener lugar en mayor proporción que cuando utilizamos el potencial basado únicamente en la topología del estado nativo, acercándonos más de esta forma al comportamiento de los sistemas reales.

5.2.1. Potencial y condiciones de simulación utilizados.

En primer lugar, hemos modificado el potencial de interacción de manera que las interacciones locales respondan a un modelo no basado en la topología del estado nativo. Las interacciones de largo alcance, sin embargo, siguen teniendo la forma de la ecuación (2.1) y vienen definidas por el mapa de contactos. De este modo podremos ver

si un tipo de interacción local más genérica favorece la formación de agregados.

La clasificación de las interacciones locales sigue siendo la misma: interacciones 1-3 (ángulo de enlace virtual) e interacciones 1-4 (torsión). Para describir las interacciones en función de la distancia entre C_α hemos calculado el histograma de distancias 1-3 y 1-4 de las proteínas de una base de datos de 1590 proteínas ya utilizada anteriormente (ver Sección 4.2 en el capítulo anterior). Para ello, hemos contabilizando el número de distancias de tipo 1-3 que están en torno a un valor d^{13} dado con respecto al número total de pares 1-3 en toda la base de datos, haciendo lo mismo con las distancias de torsión. Los histogramas de distancias 1-3 (d^{13}) y 1-4 (d^{14}) calculados se pueden ver en la Figura 5.9. En dicha figura podemos ver que el histograma para d^{13} presenta dos picos. El pico más pronunciado corresponde a las distancias d^{13} que aparecen dentro de las hélices α . El otro pico corresponde a distancias 1-3 que se producen en las hebras β . El histograma de d^{14} , sin embargo, presenta tres picos diferentes, uno de los cuales aparece a valores negativos de distancia. En el Capítulo 2 hemos explicado que las distancias 1-4 presentan, dentro de nuestro modelo, distinto signo dependiendo de su quiralidad. Los dos picos menos intensos corresponden a valores de d^{14} que ocurren dentro de las hebras β y cubren los mismos valores de distancia en valor absoluto. Debido a la disposición de los aminoácidos en zig-zag dentro de estos elementos de estructura secundaria¹ estas distancias cambian alternativamente de signo a lo largo de la secuencia de una hebra. El pico más intenso no tiene un equivalente a valores negativos ya que corresponde a distancias entre C_α que forman parte de las hélices α .

A partir de estos histogramas de distancia hemos calculado un potencial de manera similar a como se calculan los potenciales estadísticos de campo medio. Este tipo de potenciales se deducen a partir de la información que proporciona un conjunto de estructuras (que en este caso consta de 1590 proteínas) utilizando la ley de Boltzmann¹⁵⁶. Para calcular el potencial $u_{ij}(d_{ij})$ en torno a la distancia d_{ij} se utiliza una ecuación del

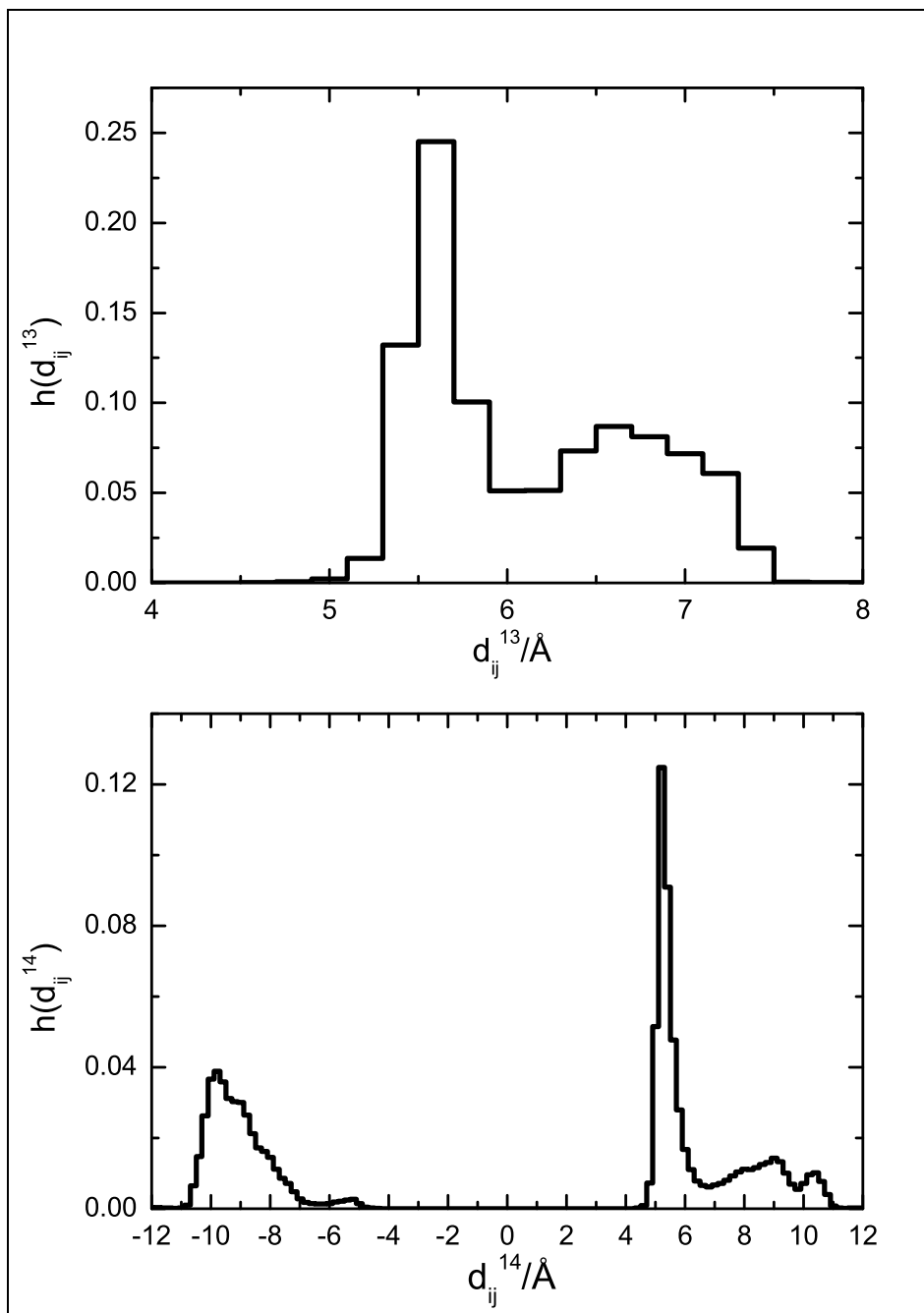


Figura 5.9: Histograma de distancias de los pares 1-3 y 1-4 de una base de datos de 1590 proteínas.

tipo:

$$u(d_{ij}) = -k_b T \ln \frac{h^{obs}(d_{ij})}{h^{ref}(d_{ij})} \quad (5.1)$$

donde k_b es la constante de Boltzmann y T es la temperatura. En esta ecuación, $h^{obs}(d_{ij})$ es el número de veces, con respecto del total, que los centros i y j aparecen a una distancia en torno a d_{ij} y $h^{ref}(d_{ij})$ indica la misma propiedad en un estado de referencia. Si utilizamos como estado de referencia un polímero totalmente flexible, $h^{ref}(d_{ij})$ se puede aproximar a una constante dentro del intervalo de distancias posibles. Además, utilizando, como hasta ahora, unidades reducidas de energía, es decir, referidas a $k_b T$, la ecuación (5.1) se puede reducir a:

$$u(d_{ij}) = -c \ln h^{obs}(d_{ij}), \quad (5.2)$$

donde c es una constante de normalización, de manera que el valor mínimo de $u_{ij}(d_{ij})$ se encuentra en -1 (como hemos hecho hasta ahora con el potencial de tipo Gō). Este tratamiento nos lleva a la obtención de las funciones de potencial para las interacciones genéricas 1-3 y 1-4 de la Figura 5.10. Dado que el histograma está discretizado, el potencial también lo está. Además, el intervalo de distancias para el cual se define el potencial corresponde al intervalo de distancias posibles, teniendo en cuenta la distancia de los enlaces virtuales entre C_α y las restricciones de volumen excluido.

Para estudiar el comportamiento de un sistema bajo este potencial de interacción, en primer lugar hemos realizado simulaciones con una única cadena de cada una de las proteínas de la sección anterior. Posteriormente, para estudiar el efecto sobre el proceso de agregación, hemos simulado sistemas de dos cadenas de las mismas proteínas. La concentración utilizada es la misma que en la sección anterior (100 mg/mL).

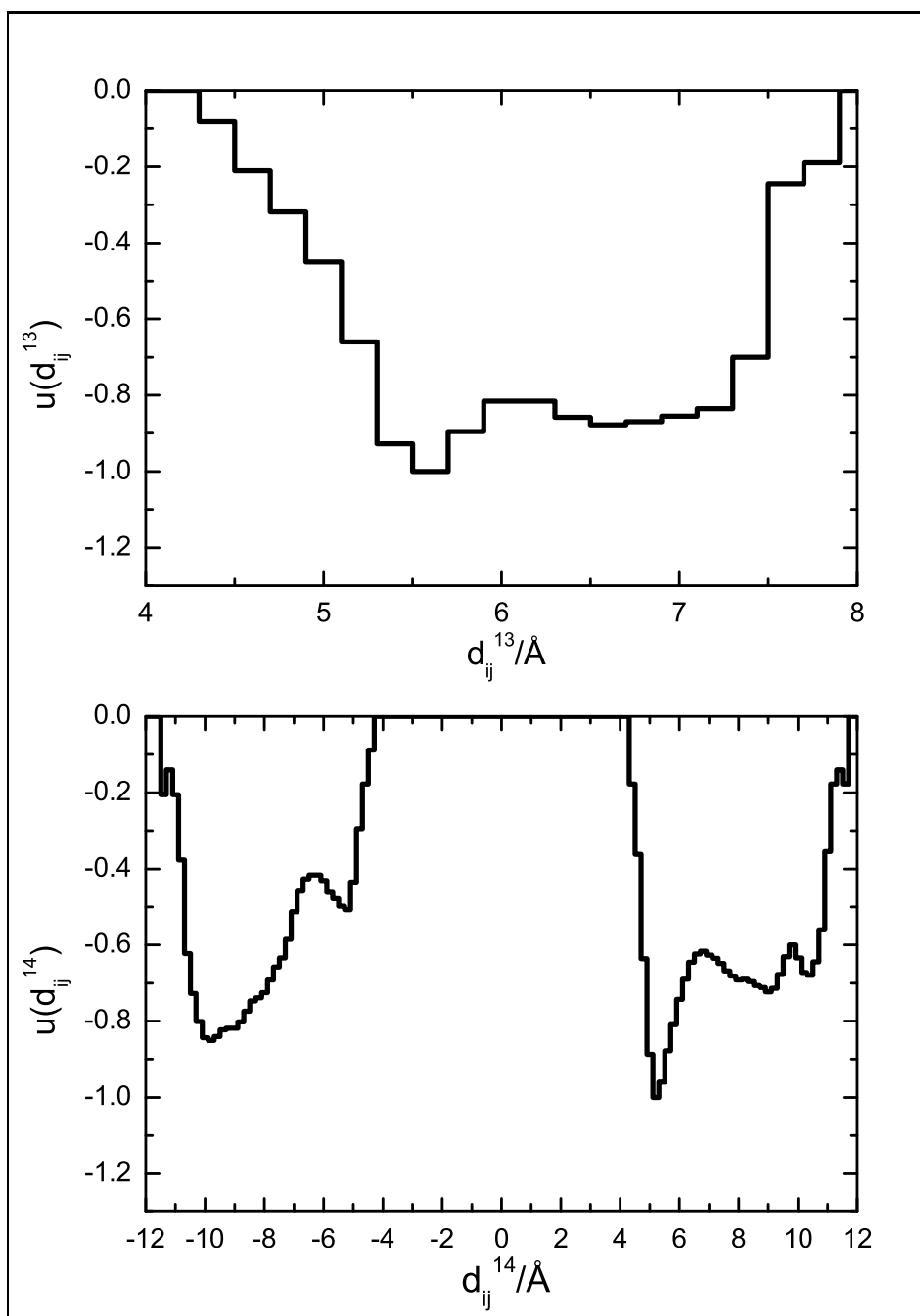


Figura 5.10: Potenciales genéricos de interacción 1-3 y 1-4.

5.2.2. Sistemas de una y dos cadenas.

Utilizando el potencial genérico de interacciones locales descrito y el potencial basado en la topología del estado nativo para las interacciones de largo alcance, hemos realizado, en primer lugar, simulaciones de sistemas de una cadena de las cuatro proteínas ya mencionadas para poder comprobar el funcionamiento y validez del potencial diseñado y, principalmente, para tener un sistema de referencia con el que evaluar el efecto de la concentración. Para poder estudiar los procesos que tienen lugar hemos calculado, como en casos anteriores, las curvas de capacidad calorífica para estos sistemas a partir de los datos de las simulaciones. Estas curvas están representadas en la Figura 5.11. Centrándonos en los sistemas de una sola cadena, lo primero que observamos es que las curvas de C_v^* de 2GB1 y 1YPB presentan dos picos. Las curvas de 1BBL y 1W4E, sin embargo, presentan un único pico. Para poder evaluar si el potencial es capaz de encontrar el estado nativo, hemos calculado el valor medio de *rmsd* respecto a la estructura nativa de las configuraciones que aparecen en nuestras simulaciones a baja temperatura. La temperatura elegida depende de la curva de capacidad calorífica. En los casos en que ésta sólo presenta un pico, hemos hecho el análisis a $T_{str} = 0.75T_m^*$, como en la Sección 5.1.2. Cuando la capacidad calorífica presenta dos picos, hemos calculado el valor de *rmsd* a $T_{str}^* = 0.75T_1^*$, siendo T_1^* y T_2^* las temperaturas menor y mayor a las que la capacidad calorífica presenta un máximo, respectivamente. Los valores obtenidos se comparan en la Tabla 5.3 con los valores de *rmsd* de las configuraciones que aparecen a una temperatura equivalente de las simulaciones realizadas con el potencial basado en la topología del estado nativo sin modificar. Los resultados de la Tabla 5.3 muestran que, aunque los valores de *rmsd* no son tan bajos como para el potencial sin modificar, en la mayoría de los casos se llega a una estructura muy parecida a la nativa. Al introducir para estas interacciones locales un potencial menos específico, estas conformaciones no están forzadas a adquirir las distancias locales que aparecen en la estructura nativa, por lo que la estructura secundaria está peor definida y el valor medio de *rmsd* aumenta. En

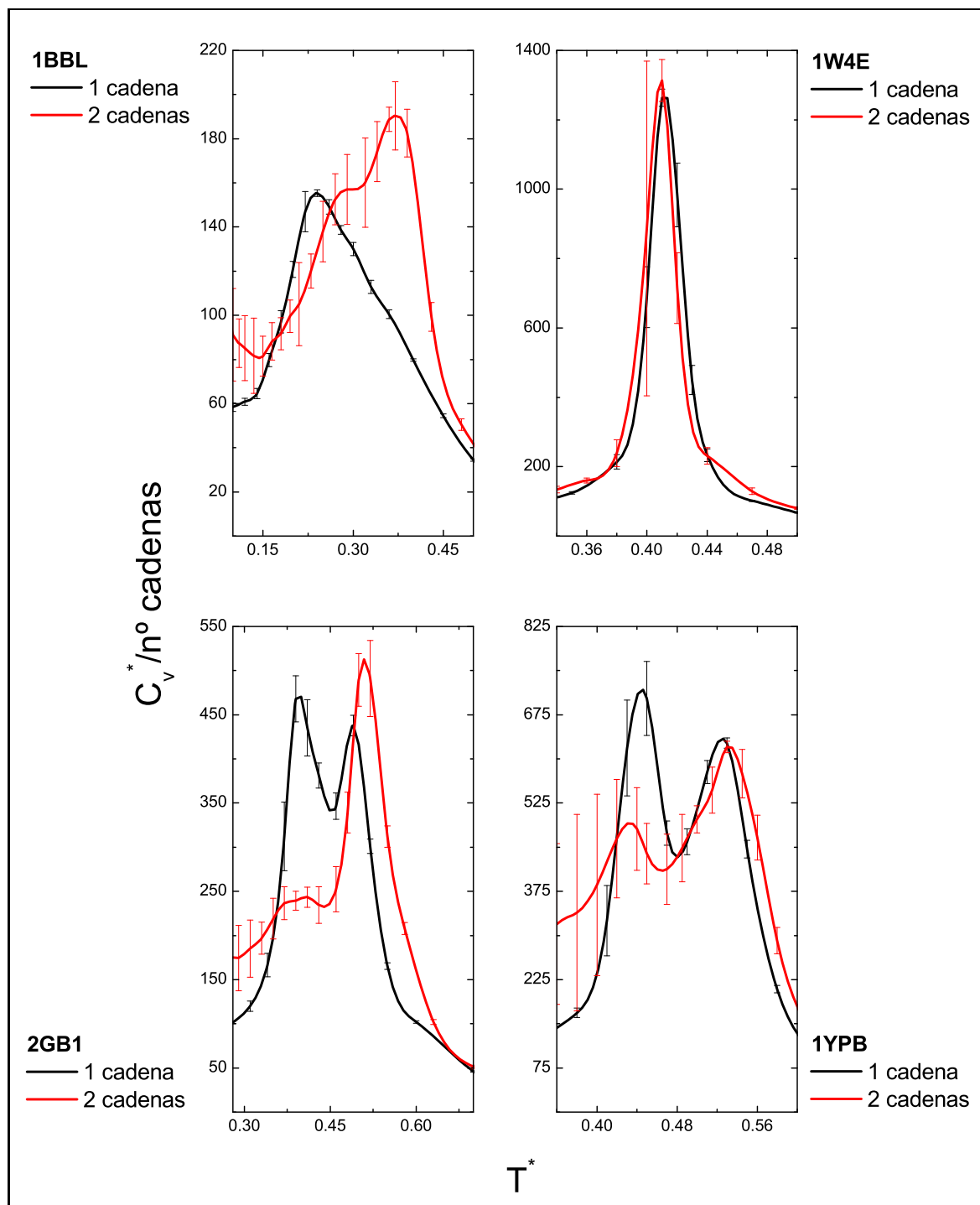


Figura 5.11: Capacidad calorífica para sistemas de una y dos cadenas de las proteínas indicadas utilizando el potencial con las interacciones locales genéricas.

Tabla 5.3: Valor medio de $rmsd$ respecto a la estructura nativa de las estructuras recogidas a una temperatura T_{str}^* (ver texto) en sistemas de una cadena para el potencial de tipo Gō sin modificar ($rmsd^G$) y para el potencial con las interacciones locales modificadas ($rmsd^L$), con la correspondiente desviación típica.

| | $rmsd^G/\text{\AA}$ | $rmsd^L/\text{\AA}$ |
|------|---------------------|---------------------|
| 1BBL | 2 ± 1 | 3.2 ± 0.9 |
| 1W4E | 1.7 ± 0.6 | 2.8 ± 0.5 |
| 2GB1 | 0.6 ± 0.1 | 1.4 ± 0.3 |
| 1YPB | 0.8 ± 0.3 | 1.5 ± 0.3 |

el caso de 1BBL y 1W4E, el valor medio de $rmsd$ es bastante elevado comparado con los valores obtenidos para las otras proteínas. Esto se debe a que, como hemos visto en el Capítulo 4, estas proteínas presentan una proporción de contactos locales mayor del 50 %. Dado que el potencial de interacciones locales es más inespecífico, la estructura secundaria está peor definida que la terciaria y esto ocasiona que la estructura global a la que se llega para estas proteínas se parezca menos a la nativa que en las demás. Aunque el valor de $rmsd$ a esa temperatura es relativamente alto, la disposición de los elementos de estructura secundaria entre sí es igual a la que aparece en la estructura nativa, dando lugar a una topología muy similar a ésta, por lo que consideramos que el potencial es capaz de encontrar el estado nativo (lo que se confirma de forma inequívoca a temperaturas inferiores).

Aunque el potencial es capaz de reproducir la estructura del estado nativo, no ocurre lo mismo con las características de la transición. Como hemos visto en la Figura 5.11, la transición de 2GB1 y 1YPB presenta dos máximos de capacidad calorífica, señal de un proceso termodinámico de tres estados. Como ya hemos hecho en capítulos anteriores de esta Tesis, para caracterizar el intermedio de plegamiento que aparece en la simulación en estos casos, hemos calculado la frecuencia con que un contacto nativo

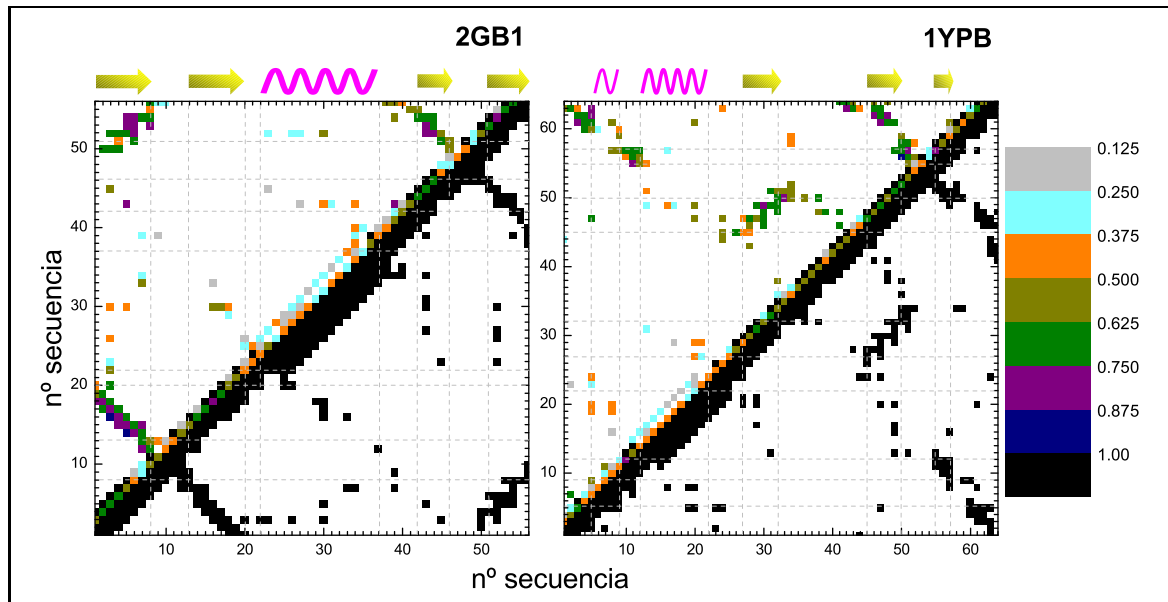


Figura 5.12: En la parte superior izquierda de los mapas se representa la frecuencia de que un contacto nativo esté formado a la temperatura a la que las curvas de capacidad calorífica a volumen constante de sistemas de una cadena de 2GB1 y 1YPB presentan un mínimo. En los triángulos inferiores derechos se indican los contactos del estado nativo.

está formado en el conjunto de configuraciones a la temperatura del mínimo entre los dos picos de capacidad calorífica. Según esta frecuencia, representada en la Figura 5.12, vemos que a la temperatura intermedia entre los dos máximos de capacidad calorífica la estructura terciaria tiende a estar mejor formada que la secundaria. Concretamente, los contactos de las hélices aparecen formados en un número muy bajo de conformaciones, mientras que la frecuencia de contacto entre los extremos de las proteínas es bastante elevada, indicando que se trata de estructuras compactas. Sin embargo, la estructura terciaria tampoco está correctamente formada, como indica la baja frecuencia del resto de contactos de largo alcance (sobre todo en el caso de 2GB1). En un proceso de enfriamiento, a mayor temperatura se produce el colapso de la cadena y la transición a menor temperatura se debe al ajuste de la estructura terciaria junto a la formación de los contactos locales, de manera que se forma el estado nativo. Si lo que se lleva a cabo es un aumento gradual de la temperatura, el proceso que tiene lugar es el inverso.

La modificación que hemos introducido al potencial confiere una mayor flexibilidad a las cadenas, ya que el potencial para las interacciones locales de la Figura 5.10 permite una variabilidad estructural mucho mayor que el potencial de la Figura 2.1. En el Capítulo 4 vimos que aumentar la flexibilidad modifica de manera sustancial la transición, dando lugar a la aparición de intermedios. El intermedio que se observa para 2GB1 para cadenas flexibles con el potencial de $G\bar{o}$ en la parte derecha de la Figura 4.2 se parece mucho al que vemos en la Figura 5.12. Además, también hemos visto en el capítulo anterior que la flexibilidad tiene un efecto adicional sobre la posible presencia de una barrera de energía libre entre el estado plegado y el desnaturalizado. Para ver si este potencial tiene el mismo efecto en el plegamiento de las cuatro proteínas estudiadas, hemos calculado los perfiles de energía libre a T_m^* o, en su caso, a T_1^* y T_2^* , para dichas proteínas. Estos perfiles se pueden ver en la Figura 5.13. En el caso de 1BBL la transición no se modifica de manera importante con respecto al potencial sin modificar. Como ya hemos visto en el capítulo anterior, la flexibilidad no afecta sustancialmente al tipo de transición en el caso de un plegamiento sin barrera. Sin embargo, en el caso del plegamiento de dos estados, una variación importante en la flexibilidad afecta a la altura de la barrera de energía libre. En el caso de 1W4E, el aumento de la flexibilidad de cadena al modificar el potencial de interacción de corto alcance no es lo suficientemente grande como para que la barrera desaparezca, pero su altura disminuye considerablemente (comparar Figuras 5.4 y 5.13). En el caso de 2GB1 y 1YPB, la barrera ha desaparecido y el plegamiento simulado de estas proteínas se produce de manera gradual. En el capítulo anterior hemos visto que la topología de 1W4E favorece la cooperatividad y que, ni siquiera haciendo modificaciones importantes en el peso de las interacciones locales y de largo alcance, desaparece totalmente la barrera de energía libre. Por eso, el aumento en la flexibilidad que hemos introducido modificando el potencial no es suficiente como para eliminar por completo la cooperatividad del tránsito.

Para evaluar el efecto de estas modificaciones en el tránsito sobre el proceso de

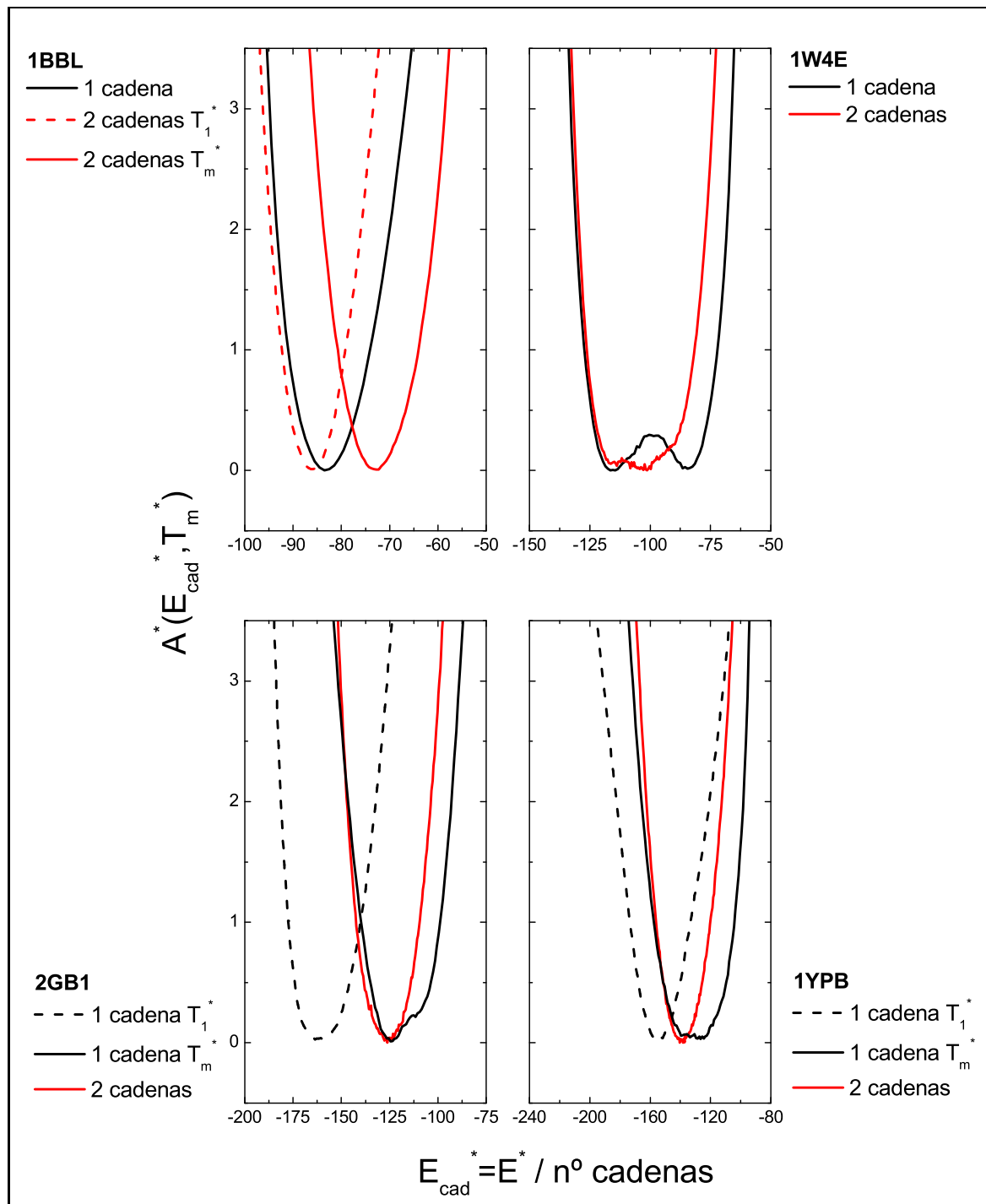


Figura 5.13: Perfiles de energía libre a T_1^* y a $T_2^* = T_m^*$ para sistemas de una y dos cadenas.

agregación, hemos realizado simulaciones de sistemas de dos cadenas. En la Figura 5.11 podemos ver que la transición sufre cambios notables con respecto al sistema de una sola cadena. La curva de capacidad calorífica de 1W4E, que tiene un plegamiento de dos estados, presenta, aproximadamente, las mismas características en el caso de dos cadenas que en el de una. Sin embargo, el pico de C_v^* que aparece a menor temperatura para 2GB1 y 1YPB reduce considerablemente su intensidad. En el caso de 2GB1, pasa a convertirse en un hombro del segundo pico más intenso. Para 1BBL se observa un comportamiento similar: la curva de capacidad calorífica presenta un pico a mayor temperatura y un hombro a temperatura aproximadamente igual a la T_m^* del sistema de una cadena.

Los cambios en las curvas de capacidad calorífica de 2GB1 y 1YPB pueden ser debidas a que los intermedios de plegamiento de los sistemas de una y dos cadenas presentan características diferentes. Debido a la presencia de otras cadenas, las interacciones intermoleculares pueden provocar variaciones en el intermedio de plegamiento y, por tanto, impedir, en cierta medida, el tránsito que tiene lugar a menor temperatura. Para corroborar esta hipótesis hemos analizado la estructura media de las configuraciones que aparecen a la temperatura del mínimo de capacidad calorífica en sistemas de dos cadenas. En la Figura 5.14 se pueden ver los mapas de frecuencia de contacto característicos de este tipo de análisis. En estas figuras, en los triángulos superiores **a** aparecen los contactos que tienen lugar dentro de cada cadena y, en el cuadrado **b** fuera de la diagonal, los que tienen lugar entre cadenas. Además, en los triángulos inferiores **c** de la diagonal se indican los contactos de las cadenas en su estado nativo, sin formar agregado. Si nos fijamos en cada cadena individualmente, vemos que la frecuencia de contacto intra-cadena es diferente de la misma frecuencia para el intermedio que aparece en las simulaciones con una sola cadena (ver Figura 5.12), sobre todo en el caso de 2GB1. El intermedio que hemos detectado con una sola cadena presenta una estructura colapsada en la que la frecuencia de los contactos entre la primera hebra β y la última es muy elevada. Sin embargo, en el caso de sistemas de dos cadenas estos mismos con-

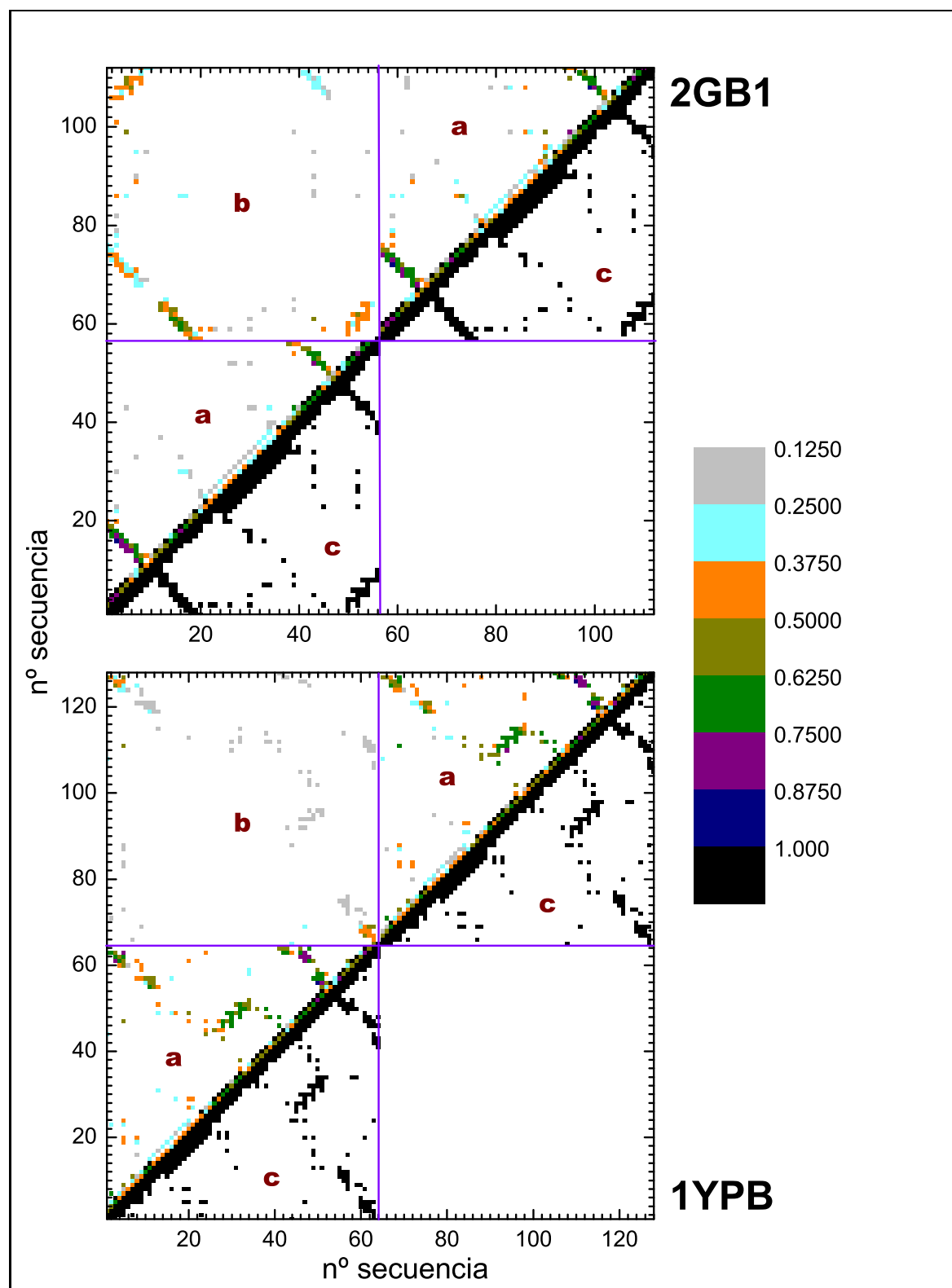


Figura 5.14: En la parte superior izquierda de los mapas se representa la frecuencia con que aparece un contacto nativo a la temperatura intermedia entre el hombro y el pico o entre los dos mínimos de las curvas de capacidad calorífica de sistemas de dos cadenas de 2GB1 y 1YPB: En los triángulos **a** se representan las interacciones intra-cadena y en el cuadrado **b** las interacciones inter-cadena. En los triángulos inferiores de la diagonal (**c**) se indican los contactos del estado nativo.

tactos ocurren con poca frecuencia dentro de cada cadena mientras que la interacción entre esas mismas hebras de moléculas diferentes está más favorecida, tal y como se puede ver a partir de la frecuencia de estos contactos en el cuadrado **b**. Esto es indicativo de un mecanismo de agregación por intercambio de dominios similar al que hemos observado para esta proteína en la sección anterior. En el caso de 1YPB, la frecuencia de contacto dentro de cada cadena no difiere tanto del intermedio de una cadena. Además, la frecuencia de los contactos inter-cadena (cuadrado **b**) es bastante baja para la mayoría de ellos, pero indica que a esa temperatura hay formados agregados además de estados en que las cadenas individuales están colapsadas. Ya hemos explicado que, debido a su simetría, 2GB1 tiene, según nuestro modelo, una mayor tendencia a formar agregados por intercambio de dominios. Por lo tanto, la presencia de estos agregados es mayor en 2GB1 que en 1YPB y por eso la transición a temperatura T_1^* se ve más afectada por el aumento de concentración en 2GB1 (ver Figura 5.11). En resumen, podemos decir que el intermedio de plegamiento está en equilibrio con agregados formados por intercambio de dominios. Esto se corresponde con estudios experimentales hechos para mutantes de la proteína GB1, que dan lugar a dímeros de este tipo a partir de intermedios de plegamiento cuya estructura es compacta aunque no del todo parecida a la nativa^{154,157}. Las mutaciones que provocan que la proteína dimerice provocan que su estructura sea menos compacta. Este efecto es comparable en nuestro modelo con un aumento en la flexibilidad, que también reduce la tendencia al colapso de la cadena. Podemos establecer así una relación entre la tendencia a la agregación y las modificaciones que provocan la aparición de intermedios de plegamiento, como la flexibilidad o las mutaciones puntuales.

Por otra parte, 1BBL presenta, en el caso de dos cadenas, un pico y un hombro en la curva de capacidad calorífica (Figura 5.11). El máximo absoluto de dicha curva aparece a temperatura mayor que el valor de T_m^* del sistema de una sola cadena. El aumento de concentración modifica por lo tanto las características de la transición.

Para ver cuál es el proceso que tiene lugar a mayor temperatura hemos analizado el conjunto de conformaciones que aparecen a la temperatura intermedia entre el pico y el hombro de C_v^* ($T^* = 0.32$) y lo comparamos con el conjunto de conformaciones que aparecen a esa misma temperatura en un sistema de una cadena. Con este objetivo, hemos calculado la frecuencia de contactos representada en la Figura 5.15 a partir de los ficheros de configuraciones recogidos a esa temperatura. Observamos que la estructura media de cada cadena aislada en el sistema concentrado es similar a la que presenta, a esa temperatura, el sistema de una cadena. Pero además, a esa temperatura aparecen contactos intermoleculares en el sistema de dos cadenas. Es decir, la transición que tiene lugar a mayor temperatura en el sistema de dos cadenas es aparentemente una transición de agregación. Más adelante, profundizando en la transición y analizando las configuraciones que se alcanzan a baja temperatura veremos que, efectivamente, se forma un agregado.

Al contrario de lo que sucede en las tres proteínas analizadas hasta el momento, en el caso de 1W4E la curva de capacidad calorífica del sistema de dos cadenas es equivalente a la de la cadena aislada (ver Figura 5.11). Podemos decir entonces que el aumento de la concentración afecta en menor medida al plegamiento de 1W4E. La transición de plegamiento es, en principio, de primer orden. Sin embargo, la barrera prácticamente desaparece al aumentar la concentración (ver Figura 5.13). Este efecto ya lo hemos observado para sistemas de cuatro cadenas de 2GB1 cuando utilizamos el potencial de Gō cruzado. Más adelante veremos que, debido a la pequeña barrera de energía libre, se produce una estabilización de estados intermedios que da lugar a una modificación de las características del tránsito.

Para poder profundizar en las características de la transición y estudiar con más detenimiento cómo intervienen las interacciones inter e intramoleculares, hemos calculado también, utilizando el método WHAM, los mapas bidimensionales de energía libre. Como ya hemos dicho, las variables más adecuadas para este objetivo son los contactos

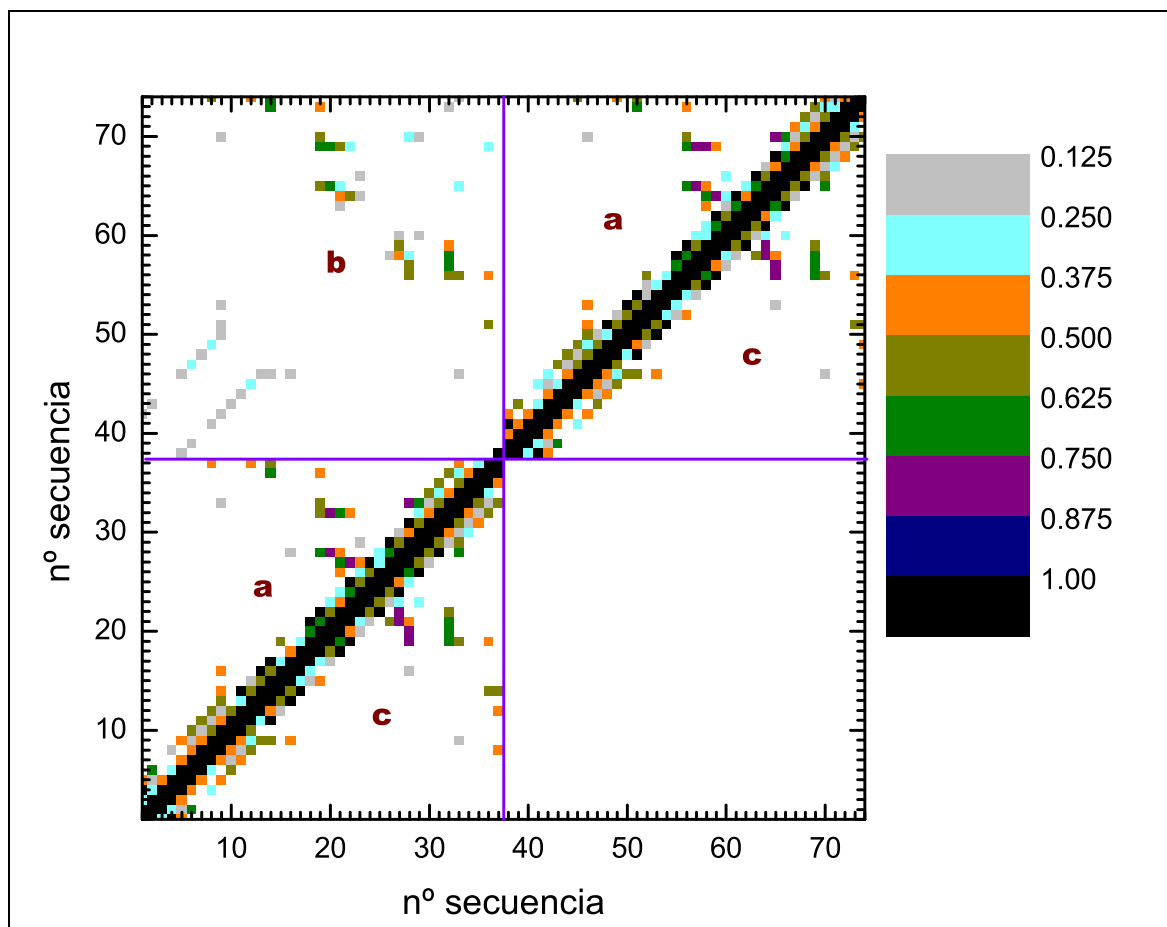


Figura 5.15: Frecuencia con que aparece un contacto nativo a $T^* = 0.32$ para los sistemas de una y dos cadenas de 1BBL. La parte superior izquierda corresponde a sistemas de dos cadenas: En los triángulos **a** se representa la interacción intra-cadena y en el cuadrado **b** la interacción inter-cadena. Los triángulos inferiores **c** de la diagonal corresponden a las interacciones intra-cadena de sistemas de una cadena a esa misma temperatura.

intra- e inter-cadena. Así podemos hacernos una idea de qué grado de estructura está formada y hasta qué punto las cadenas están interaccionando entre sí a la temperatura de estudio. En la Figura 5.16 hemos representado estos mapas para las cuatro proteínas que estamos estudiando, junto con los histogramas de contactos intramoleculares a la misma temperatura en los sistemas de una cadena.

En 1BBL, a T_m^* las cadenas adquieren un grado de estructura terciaria similar al que presentan en el sistema de una sola cadena. Pero además, se produce un importante número de contactos inter-cadena, lo cual da lugar a una transición a esta temperatura

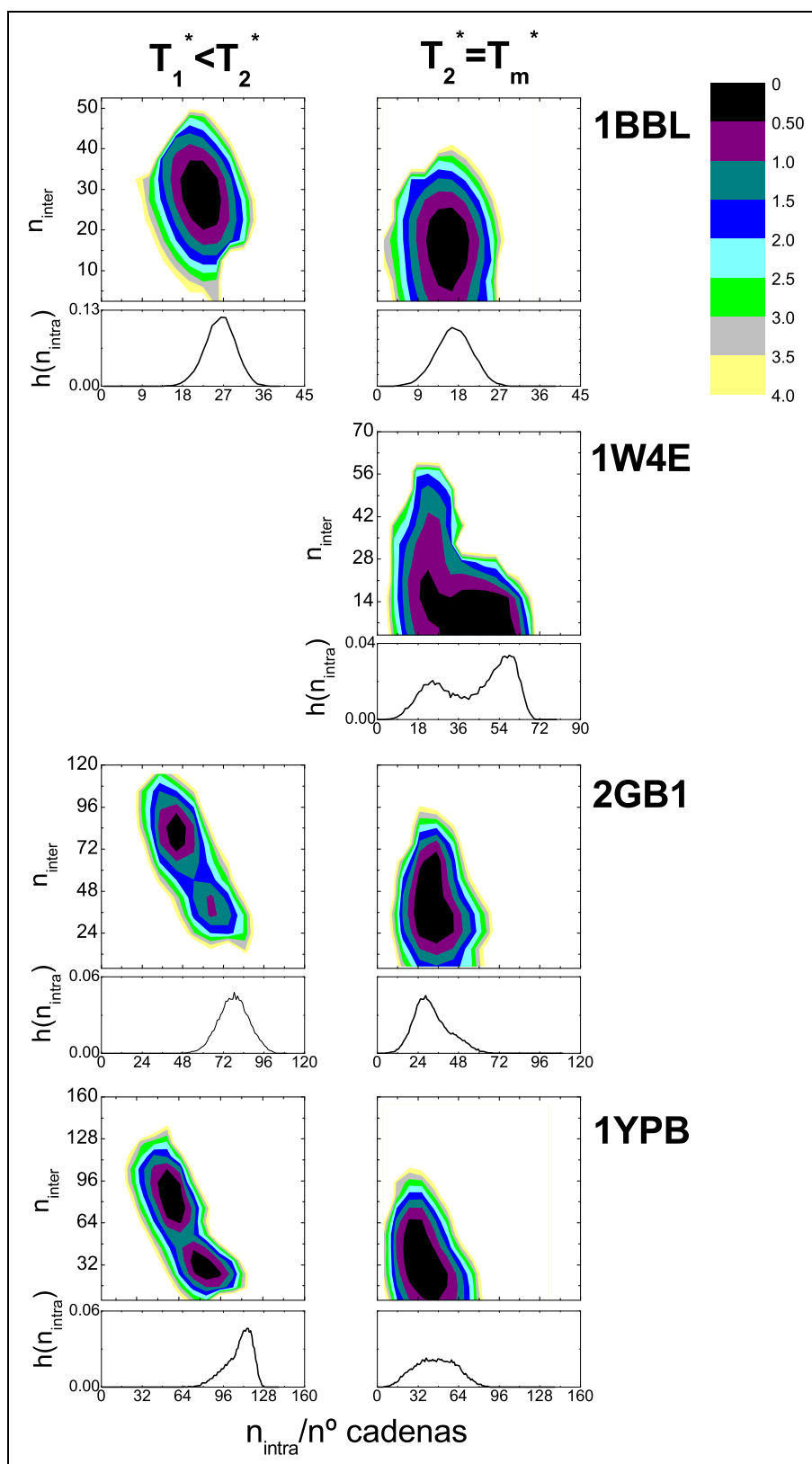


Figura 5.16: Mapas de energía libre en función de los contactos de largo alcance inter- e intra-cadena a las temperaturas a las que aparece un máximo de capacidad calorífica. Debajo de cada mapa aparece el histograma de contactos intra-cadena de los sistemas de una cadena a la misma temperatura.

que no tiene lugar en sistemas diluidos. Es decir, a esta temperatura tiene lugar un proceso de agregación de las cadenas en el estado desnaturalizado. Como hemos visto en la Figura 5.15, esto da lugar a un intermedio en que las cadenas tienen parte de la estructura nativa formada pero, además, interaccionan entre ellas. Al disminuir la temperatura, la proteína adquiere la mayoría de los contactos de la estructura terciaria nativa, casi igual que en el caso de una cadena, pero además se forma un mayor número de interacciones intermoleculares. El proceso de agregación en este caso está favorecido por dos motivos: por un lado, la ausencia de barrera de energía libre permite que haya interacción entre conformaciones de cadena intermedias entre el estado plegado y el desnaturalizado; por otra parte, debido a la flexibilidad del sistema y a la estructura de 1BBL, es posible que las cadenas formen agregados a la vez que mantienen un elevado grado de estructura nativa. Si bien estos agregados son similares a los estados asociados que observamos con el potencial original, el grado de estructura nativa, en este caso, es menor. Si comparamos los mapas de energía libre a menor temperatura de esta proteína en las Figuras 5.5 y 5.16 vemos que el número de interacciones inter-cadena es mayor y el de interacciones intra-cadena menor en ésta última. Es decir, se pierde estructura nativa a favor de una mayor interacción intermolecular.

1W4E es un homólogo estructural de 1BBL; sin embargo, su estructura es mucho más compacta y la topología de contactos de largo alcance, más compleja. Debido a esto, según nuestro modelo, esta proteína pliega a través de una transición de dos estados, hasta cierto punto incluso en el caso de la importante flexibilidad que supone introducir un potencial genérico para las interacciones locales. La pequeña barrera de energía libre que aparece entre el estado plegado y el desnaturalizado (ver Figura 5.13) hace que el proceso de agregación esté más impedido que en el caso de 1BBL. Sin embargo, ya hemos visto que, incluso en estos casos, siempre existe cierta competencia entre este proceso y el plegamiento. Además, con el modelo de interacciones que utilizamos en esta sección, la barrera entre el estado plegado y el desnaturalizado es bastante baja. La población de

estados intermedios es, por lo tanto, lo suficientemente alta como para que exista cierta interacción entre ellos, de manera que se produce su estabilización por la formación de contactos intermoleculares, al aumentar el número de cadenas (ver Figura 5.16). Como consecuencia, la barrera de energía libre desaparece del todo cuando la concentración es mayor, como hemos visto en la Figura 5.13.

Para 2GB1 y 1YPB observamos comportamientos similares: a $T_2^* = T_m^*$ se produce el proceso de compactación de la proteína. Este proceso, con el potencial de interacciones utilizado en esta sección, tiene lugar de manera gradual lo cual permite que, a esta temperatura, se forme un elevado número de contactos intermoleculares a la vez que se forma la estructura terciaria. A menor temperatura observamos dos mínimos de energía libre y ninguno de los dos corresponde a conformaciones con un número de contactos intra-cadena comparable al sistema de una cadena aislada. A partir de los ficheros de coordenadas resultantes de nuestras simulaciones podemos determinar que el estado que presenta un mayor número de contactos intra-cadena corresponde a cadenas colapsadas sobre sí mismas. El número de contactos no es equivalente al del sistema de una cadena porque, debido a su flexibilidad, adoptan conformaciones que permiten un mayor número de contactos inter-cadena. Por otra parte, el estado que presenta un elevado número de contactos intermoleculares está formado por configuraciones en que las cadenas están interpenetradas unas con otras. El conjunto de configuraciones es muy amplio, incluyendo conformaciones que se producen por intercambio de dominios. Pero, en la mayoría de los casos, se trata de agregados inespecíficos en que las cadenas presentan poca estructura nativa.

Las modificaciones introducidas en el potencial provocan un aumento en la flexibilidad de las cadenas que da lugar a una fuerte competición entre agregación y plegamiento. Los mínimos de energía libre de los estados agregados de la Figura 5.16 tienen una profundidad (baja energía libre) comparable con la de los mínimos que aparecen para valores mayores de contactos intra-cadena. Esto sucede incluso en el caso de 1W4E,

Tabla 5.4: Valor medio del radio de giro (R_g) y de $rmsd$ respecto a la estructura nativa (con la correspondiente desviación típica) de las configuraciones recogidas a una temperatura $T_{str}^* = 0.75T_1^*$, siendo T_1^* la menor temperatura a la que C_v^* presenta un pico o un hombro, en sistemas de dos cadenas para el potencial con las interacciones locales modificadas. R_g^{PDB} es el valor del radio de giro de la estructura nativa.

| | $R_g^{PDB}/\text{\AA}$ | $R_g/\text{\AA}$ | $rmsd/\text{\AA}$ |
|------|------------------------|------------------|-------------------|
| 1BBL | 9.3 | 11.0 ± 0.9 | 6.8 ± 1.1 |
| 1W4E | 9.0 | 9.3 ± 0.5 | 3.5 ± 1.2 |
| 2GB1 | 10.1 | 13 ± 5 | 8 ± 7 |
| 1YPB | 10.8 | 10.9 ± 0.8 | 4 ± 2 |

que presenta una transición de primer orden. ¿Quiere esto decir que el estado de mínima energía al que se llega en sistemas de varias cadenas es un agregado? Para poder dar respuesta a esta pregunta hemos calculado, al igual que hicimos en los sistemas de una cadena, el valor medio del radio de giro y del valor de $rmsd$ respecto a la estructura nativa a temperatura T_{str}^* . Los resultados obtenidos se encuentran en la Tabla 5.4. En todos los casos, los valores medios de $rmsd$ y sus desviaciones son relativamente elevados, indicando la posibilidad de que las conformaciones que aparecen a esa temperatura no correspondan al estado nativo. En algunos casos, además, el valor de radio de giro no se corresponde con el nativo.

Para poder determinar a qué se deben las desviaciones respecto del estado nativo hemos calculado la distribución de radios de giro. Esta distribución, que aparece en la Figura 5.17, la hemos calculado a la temperatura T_{str}^* en función de la energía total del sistema con el objetivo de comprobar si las diferentes configuraciones que puedan aparecer tienen igual o diferente energía. Los resultados de radio de giro presentados corresponden a una de las dos cadenas, pero hemos comprobado que son equivalentes para la otra. En dicha figura, en primer lugar vemos la distribución correspondiente a

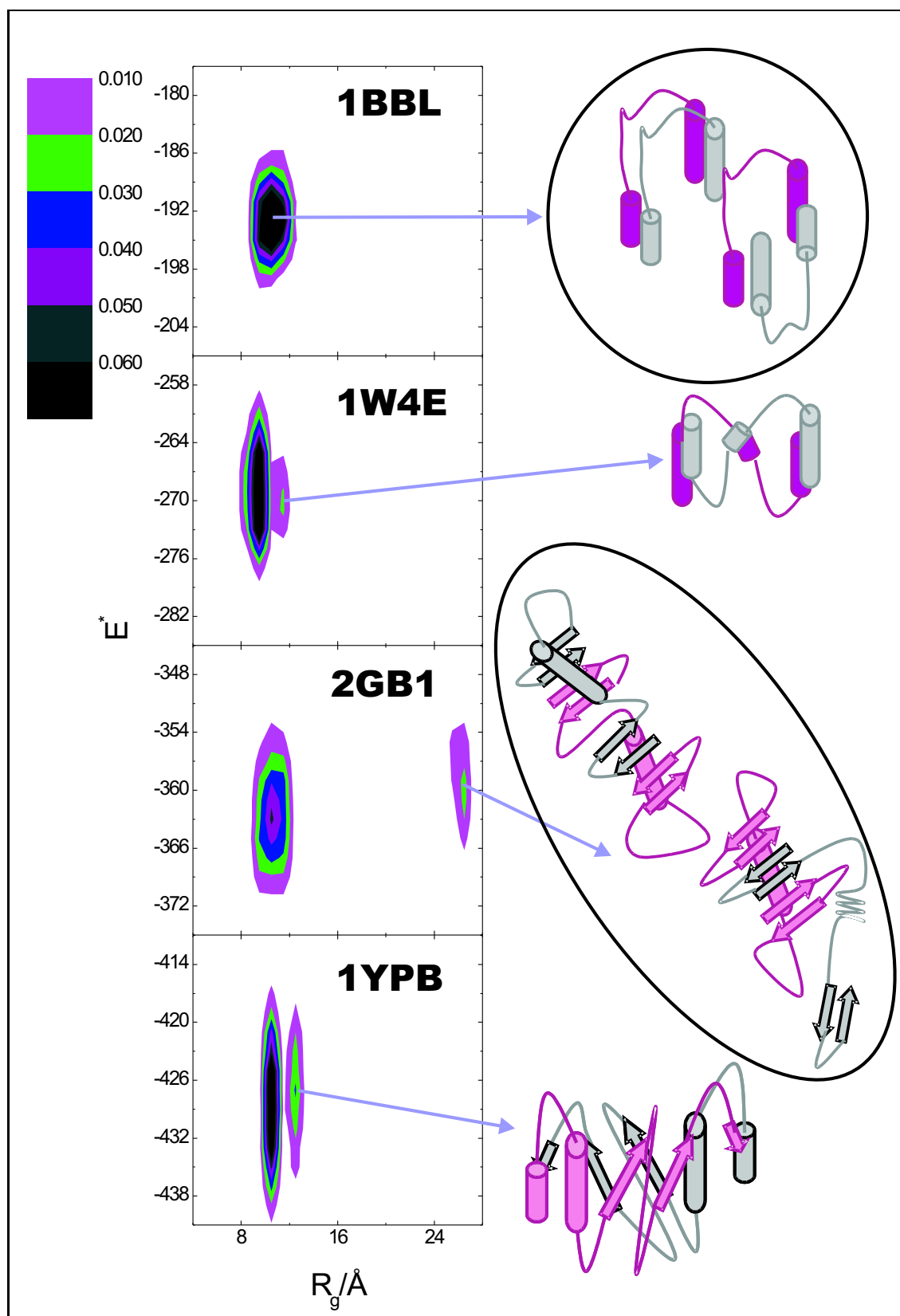


Figura 5.17: Histograma del radio de giro en función de la energía total del sistema de las proteínas indicadas a su temperatura T_{str}^* .

esa temperatura para 1BBL. Esta proteína presenta un único máximo en torno a valores del radio de giro superiores al del estado nativo. En los archivos de coordenadas hemos podido ver que estas conformaciones corresponden a las cadenas formando agregados como los que se representan esquemáticamente en la misma figura. En estos agregados la estructura secundaria de cada cadena está pobremente formada y la conformación de cada una de ellas es abierta con respecto de la disposición de las hélices en el estado nativo, dando lugar a un valor elevado de *rmsd*. Su disposición abierta permite que las cadenas puedan asociarse de dos maneras diferentes con la misma energía: una en que se produce una especie de intercambio de dominios y otra en que las cadenas se disponen paralelamente (de manera que entre ellas se producen contactos de tipo 1-5') (ver Figura 5.17). Queda clara, por tanto, la tendencia a la agregación de esta proteína bajo las condiciones de este modelo.

Por otra parte, 1W4E, 2GB1 y 1YPB presentan, a esta temperatura, dos máximos en la distribución de radios de giro que corresponden a estados degenerados en energía: uno cercano al valor del radio de giro de la estructura nativa y otro a valores mayores (ver Figura 5.17). Estos estados corresponden, para los valores menores de R_g , a configuraciones en que las dos cadenas están colapsadas, pero con una estructura terciaria no muy bien definida. Como hemos comentado, debido a su flexibilidad, las cadenas adoptan estructuras menos parecidas a la nativa para maximizar el número de contactos inter-cadena. Por otro lado, aparece un máximo de probabilidad para valores de radio de giro mayores al nativo que corresponden a un agregado formado, en el caso de 2GB1 y 1YPB, por intercambio de dominios, cuyas estructuras aparecen representadas, esquemáticamente, en la Figura 5.17. Para 1W4E el tipo de agregado que se forma es similar al que tiene lugar en 1BBL: la flexibilidad de cadena permite que éstas adopten configuraciones abiertas que pueden interaccionar entre sí (ver Figura 5.17). Podemos decir entonces que la flexibilidad de cadena supone un grado de frustración elevado que favorece enormemente el proceso de agregación, aunque con este potencial el colapso

de las cadenas sobre sí mismas sigue siendo el proceso mayoritario. La deformación con respecto a la estructura nativa de los estados colapsados y la presencia de agregados hace que el valor medio de *rmsd* sea alto en todos los casos.

Cuando modificamos el potencial de interacciones locales el equilibrio entre agregación y plegamiento se desplaza hacia la formación de agregados con respecto al potencial original. Pero, además, es importante destacar que la formación de estos agregados no siempre tiene lugar a través de un intercambio de dominios. Este mecanismo tiene lugar en el caso de 2GB1 y 1YPB, para los que ya hemos observado que es posible con el potencial no modificado. Pero en el caso de 1BBL y 1W4E, la mayor flexibilidad de las cadenas hace que éstas adopten configuraciones más abiertas que el estado nativo que favorecen la agregación de estados no colapsados.

5.3. Modificación de las interacciones de largo alcance.

En el caso anterior hemos observado cómo la modificación de las interacciones locales, de manera que el esquema de interacción sea más genérico, provoca, como cabía esperar, que la agregación esté más favorecida que en un sistema mínimamente frustrado. En este apartado estudiamos cuál es la competencia entre los procesos de agregación y plegamiento cuando modificamos el potencial basado en la topología del estado nativo de manera que sean las interacciones de largo alcance las que tengan un carácter más genérico. Introducimos así, de forma aproximada, el efecto de la secuencia. Esto nos permite estudiar otro tipo de frustración del sistema diferente y su influencia en el equilibrio entre agregación y plegamiento.

5.3.1. Potencial y condiciones de simulación utilizados.

En este caso hemos mantenido el potencial de la ecuación (2.1) para las interacciones de tipo local. Sin embargo, para las interacciones de largo alcance hemos introducido

un término adicional:

$$u_{ij}(r_{ij}) = \begin{cases} w_G^{lr} \times [(r_{ij} - d_{ij}^{nat})^2 - a^2] + w_Z^{lr} \times u_Z(d_{ij}) & ; \quad d_{ij}^{nat} - a < r_{ij} < d_{ij}^{nat} + a \\ w_Z^{lr} \times u_Z(d_{ij}) & ; \quad \text{en el resto de los casos} \end{cases} \quad (5.3)$$

En esta ecuación, $u_Z(d_{ij})$ corresponde al potencial de interacción definido por Zhou y colaboradores¹⁵⁸. Se trata de un potencial estadístico de colapso hidrófobo basado en la ecuación (5.1) que tiene en cuenta la naturaleza de los aminoácidos. En este caso el estado de referencia que utilizan los autores es el llamado estado DFIRE (*Distance-scaled, Finite Ideal-gas Reference state*). En esta Tesis hemos utilizado la deducción de este potencial realizada por Larriva y colaboradores¹⁵⁹ para la interacción entre C_α . En dicho trabajo, los autores ponen de manifiesto la baja especificidad de este potencial, que es capaz de reproducir el colapso hidrófobo de manera muy general. Debido a esto no podemos definir un potencial totalmente genérico para las interacciones terciarias y lo hemos combinado con el potencial basado en la topología del estado nativo. Los parámetros w_G^{lr} y w_Z^{lr} de la ecuación (5.3) controlan la intensidad relativa de un tipo de interacción con respecto de otro. Los valores que hemos dado a estos parámetros son 0.9 y 0.1, respectivamente, de manera que reducimos la interacción de tipo Gō al 90 % y el resto de la interacción se debe al potencial genérico que considera la secuencia de forma aproximada.

Como en el caso de la modificación de las interacciones locales, hemos simulado sistemas de una y dos cadenas, utilizando las mismas proteínas y en la misma concentración (100 mg/mL).

5.3.2. Sistemas de una y dos cadenas.

Como en la sección anterior, hemos realizado varias simulaciones tanto de sistemas de una como de dos cadenas. De esta manera podemos estudiar el efecto de la secuencia en la propia transición de plegamiento para, posteriormente, analizar su influencia

Tabla 5.5: Valores de la temperatura de transición T_m^* , deducida a partir del máximo de capacidad calorífica para sistemas de una y dos cadenas con el potencial de largo alcance modificado. El error es de la unidad en la última cifra.

| | 1 cadena | 2cadenas |
|------|----------|----------|
| 1BBL | 0.48 | 0.47 |
| 1W4E | 0.61 | 0.61 |
| 2GB1 | 0.62 | 0.61 |
| 1YPB | 0.66 | 0.655 |

en el proceso de agregación. Tomando como referencia los resultados obtenidos con el potencial basado sólo en la topología, hemos estudiado, en primer lugar, los sistemas de una cadena. Para poder establecer una comparación inicial, hemos calculado, a partir de los datos de la simulación, las curvas de capacidad calorífica a volumen constante de la Figura 5.18. Si comparamos los resultados con las curvas de la Figura 5.3, se puede observar que la transición de plegamiento de una cadena se produce a lo largo de un intervalo de temperaturas mayor en el caso del potencial modificado según lo explicado en la Sección 5.3.1. Sin embargo, la temperatura de transición con este modelo de interacción, que podemos ver en la Tabla 5.5, no varía sustancialmente respecto a las del potencial de Gō (Tabla 5.1). Podemos decir entonces que, aunque modificar las interacciones de largo alcance, en la medida relativamente pequeña en que lo hemos hecho, no afecta de manera importante a la estabilidad del modelo, sí afecta a las características de la transición que se define para el plegamiento simulado de estas proteínas. La mayor anchura de las curvas de capacidad calorífica es indicativa, como ya hemos visto en otros casos, de una disminución (o incluso desaparición) de la barrera de energía libre entre el estado plegado y el desnaturalizado. Dado que 1BBL no presenta esta barrera con el potencial basado únicamente en la topología del estado nativo, la modificación introducida no le afecta de la misma manera que a las otras tres proteínas y, por eso,

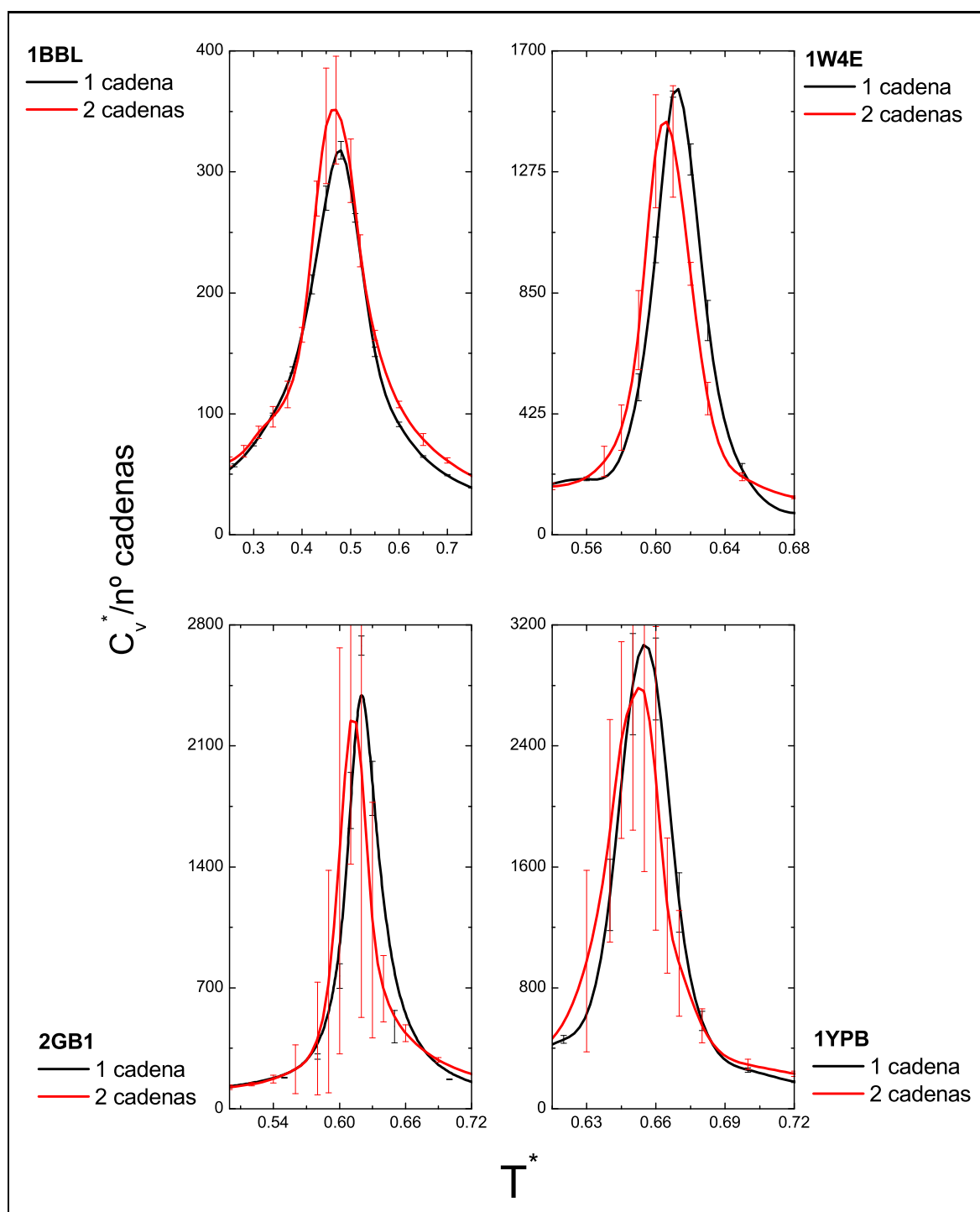


Figura 5.18: Capacidad calorífica para sistemas de una y dos cadenas de las proteínas indicadas con el potencial en el que las interacciones de largo alcance están modificadas.

Tabla 5.6: Valor medio de $rmsd$ respecto a la estructura nativa de las configuraciones recogidas a una temperatura $T_{str}^* = 0.75T_m^*$ para los sistemas de una cadena.

| $rmsd/\text{\AA}$ | |
|-------------------|---------------|
| 1BBL | 2 ± 1 |
| 1W4E | 1.8 ± 0.6 |
| 2GB1 | 0.7 ± 0.2 |
| 1YPB | 0.8 ± 0.3 |

la curva de capacidad calorífica no presenta propiedades diferentes. Con el objetivo de comprobar si, verdaderamente, los cambios producidos en la capacidad calorífica se deben a que la transición es menos cooperativa con este nuevo potencial, hemos calculado los perfiles de energía libre a la temperatura de transición. En dichos perfiles, representados en la Figura 5.19, podemos ver que la barrera es, en todos los casos, menor que cuando utilizamos el potencial de la Sección 5.1 (ver Figura 5.4). Las propiedades de la transición se han modificado de manera notable, aunque en general ocurre lo mismo que con el potencial original: 1BBL presenta un plegamiento sin barrera mientras que 1W4E, 2GB1 y 1YPB presentan, en principio, transiciones de dos estados.

Por otra parte, para evaluar la validez de este potencial, debemos saber si es capaz de encontrar el estado nativo a bajas temperaturas. Para ello hemos calculado el valor medio de $rmsd$ respecto a la estructura nativa a una temperatura $T_{str}^* = 0.75T_m^*$, como hemos hecho hasta el momento. Los resultados obtenidos aparecen en la Tabla 5.6. Comparándolos con los de la Tabla 5.2 vemos que son muy parecidos. Es decir, el potencial que estamos utilizando es capaz de llegar al estado nativo de la misma manera que el potencial basado únicamente en la topología. Por lo tanto, si en un sistema de dos cadenas se llega a otros valores diferentes, es consecuencia de un posible proceso de agregación, no de que el potencial en sí no sea capaz de llegar al estado nativo (algo que sí ocurriría si aumentáramos la contribución del potencial de Zhou).

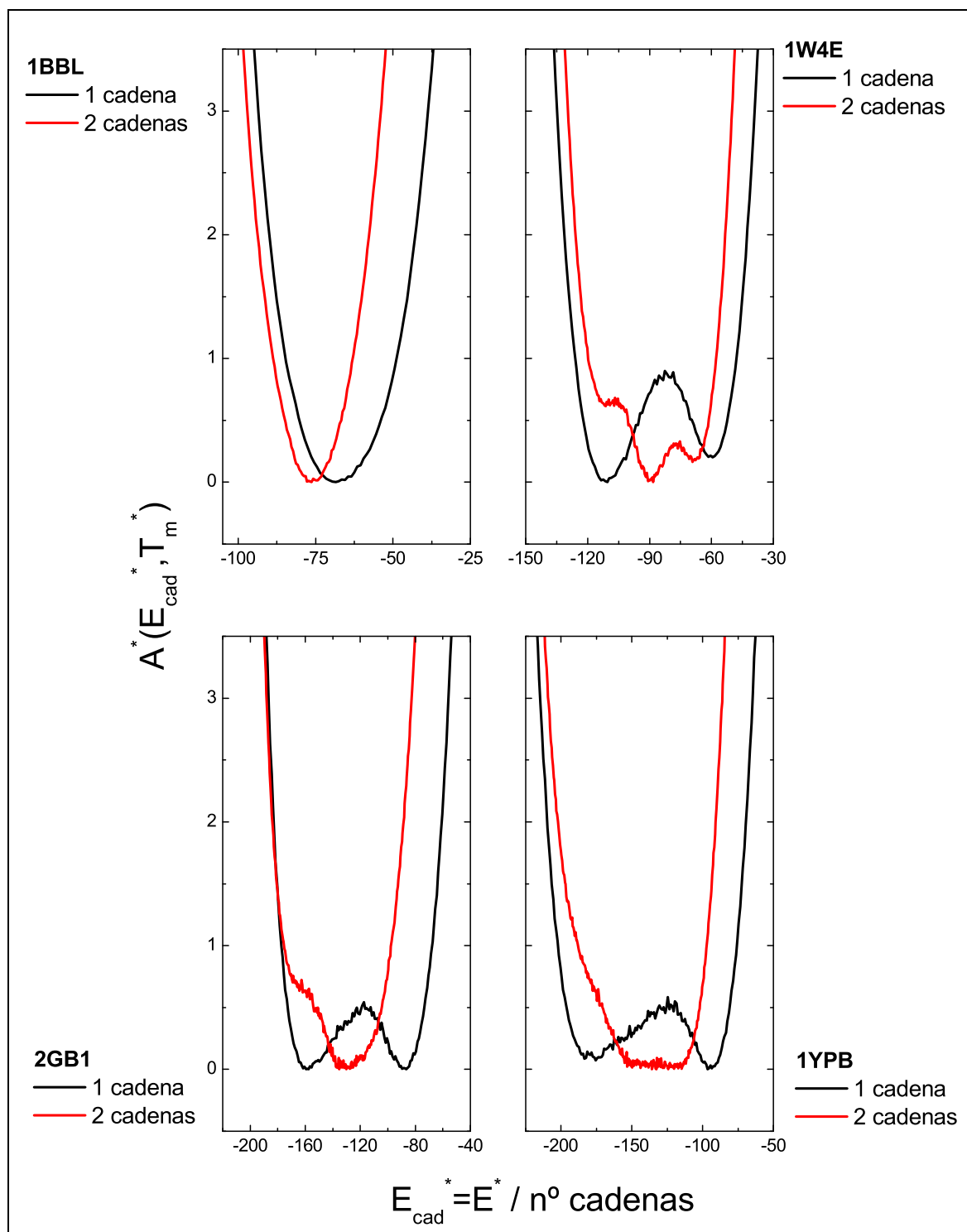


Figura 5.19: Perfiles de energía libre a T_m^* para sistemas de una y dos cadenas de las proteínas indicadas con el potencial en el que las interacciones de largo alcance están modificadas.

Para poder comprobar si la agregación se ve favorecida al modificar las interacciones de largo alcance hemos realizado simulaciones de sistemas de dos cadenas. Hemos observado que para 1BBL y 1W4E las propiedades de la transición no se modifican de manera importante, como indican las curvas de capacidad calorífica y los perfiles de energía libre de la Figuras 5.18 y 5.19, respectivamente. Sin embargo, en los casos de 2GB1 y 1YPB se produce la desaparición de la barrera de energía libre al aumentar la concentración. Para comprobar que, efectivamente, las transiciones que tienen lugar en los sistemas de dos cadenas de 2GB1 y 1YPB son procesos sin barrera, hemos calculado los perfiles de energía libre a temperaturas superiores e inferiores a T_m^* . Estos perfiles están representados en la Figura 5.20, donde podemos ver que a $T_u^* > T_m^*$ aparece un mínimo a valores poco negativos de energía, correspondiente al estado desnaturado. Este mínimo se desplaza a valores más negativos a medida que disminuye la temperatura. A T_m^* este mínimo es bastante ancho, englobando un conjunto amplio de conformaciones. Si seguimos bajando la temperatura, vemos que el perfil de energía libre presenta un hombro (en el caso de 1YPB) o una barrera muy pequeña (2GB1) entre configuraciones con energía intermedia y muy baja. La barrera de energía libre, por lo tanto, prácticamente desaparece al aumentar el número de cadenas, dando lugar a una transición diferente.

La presencia de varias cadenas en el medio provoca que la barrera de 2GB1 y 1YPB prácticamente desaparezca. Por lo tanto, la desaparición de dicha barrera debe ser una consecuencia de la competición que existe entre las interacciones inter- e intra-cadena. Para corroborar esta hipótesis es necesario profundizar en el tránsito, estudiando la posible competición entre las interacciones inter- e intra-cadena. Ya hemos comprobado cómo la mejor manera de hacerlo es a través de los mapas de energía libre en función del número de contactos inter e intramoleculares formados a la temperatura de transición. Éstos aparecen en la Figura 5.21. En estos mapas vemos que, efectivamente, los tres mínimos de energía libre que suelen aparecer (correspondientes al estado desnaturado,

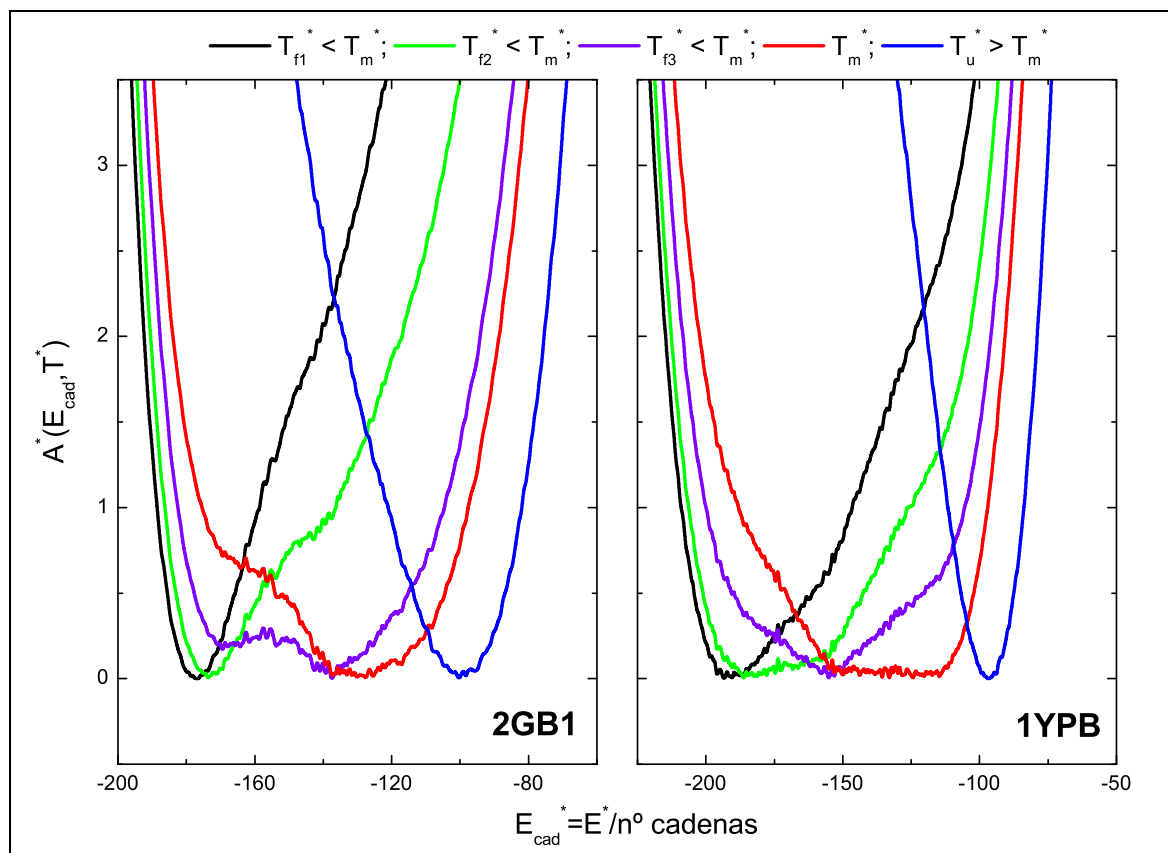


Figura 5.20: Perfiles de energía libre a cinco temperaturas diferentes para sistemas de dos cadenas de las proteínas indicadas con el potencial en el que las interacciones de largo alcance están modificadas.

do y los que corresponden al plegamiento de una y dos cadenas) aparecen solapados, especialmente en el caso de 2GB1. El conjunto de configuraciones con bajo número de contactos intra-cadena no se diferencia claramente del estado en el que el número de contactos intermoleculares es elevado. De hecho, hay un único mínimo de energía libre para valores bajos de número de contactos intra-cadena en el que el valor de contactos inter-cadena puede variar mucho (entre 20 y 60). Si analizamos los ficheros de coordenadas de las configuraciones generadas a esa temperatura a lo largo de la simulación observamos que las configuraciones que corresponden a cada zona del mapa de 2GB1 son las siguientes:

- Por un lado aparecen conformaciones con poca estructura terciaria que interaccio-

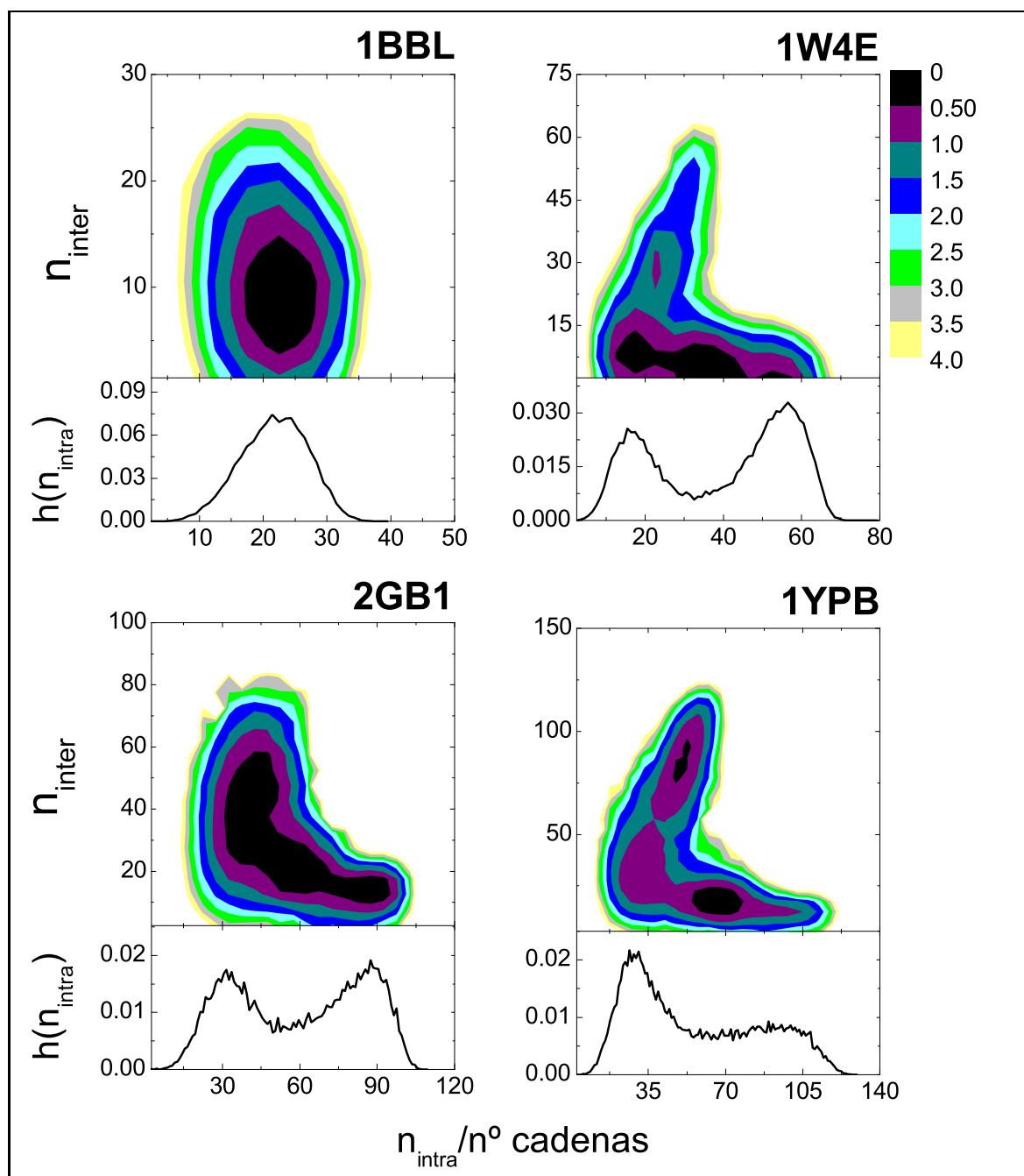


Figura 5.21: Mapas de energía libre para los sistemas de 2 cadenas de las proteínas estudiadas en función de los contactos inter e intramoleculares formados a la temperatura de transición. En la parte inferior de cada mapa se representan los histogramas de los contactos terciarios intramoleculares que están formados a esas temperaturas en los sistemas de una cadena correspondientes.

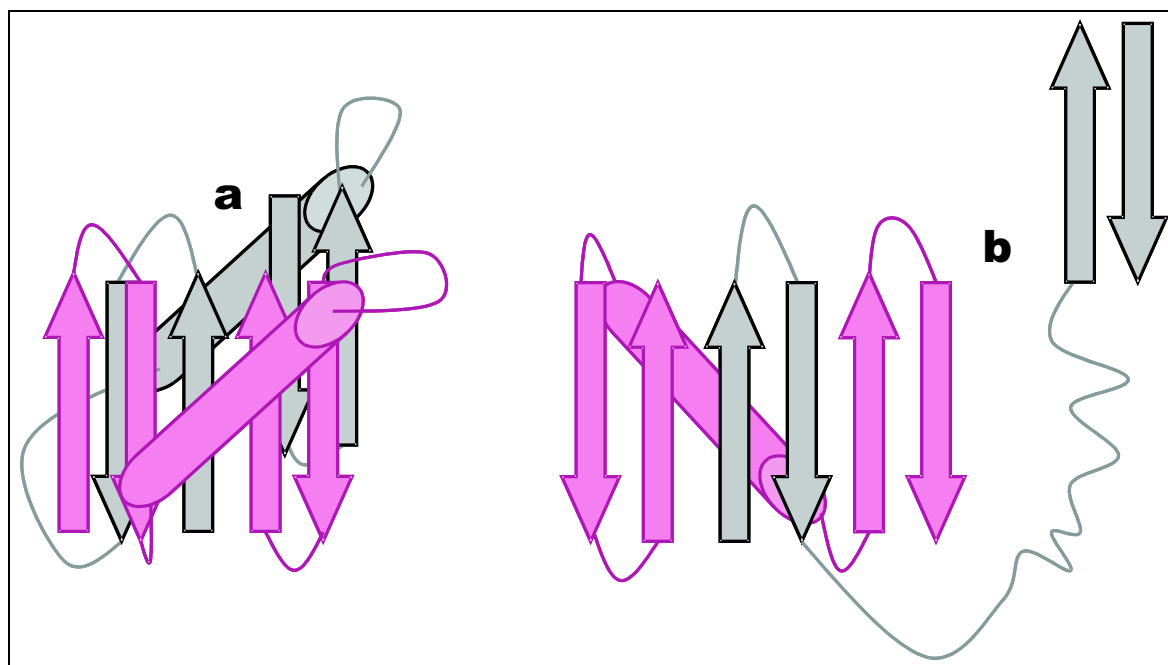


Figura 5.22: Representación esquemática de las conformaciones que dan lugar a agregados por intercambio de dominios en 2GB1.

nan más o menos entre sí. Esto da lugar a una zona en el mapa de baja energía libre con pocos contactos intra-cadena pero que tiene bastantes contactos inter-cadena. De hecho, puede dar lugar a agregados que se forman a través de un intercambio de dominios en el que lo que se intercambia es una de las hebras (como se muestra en la Figura 5.22a) o una de las horquillas (Figura 5.22b). La estructura nativa está peor formada de lo que, *a priori*, se puede entender a partir de las representaciones de la Figura 5.22. Sin embargo, es importante destacar que esta zona del mapa corresponde a configuraciones en que cada cadena tiene una media de entre 25 y 45 contactos nativos intramoleculares, lo cual indica el grado de estructura terciaria adquirido. Además, cuando se forman los contactos inter-cadena, aumenta el grado de contactos intra-cadena formados, como indica la ligera inclinación del mínimo de energía libre hacia la derecha.

- A valores mayores de número de contactos intra-cadena aparecen conformaciones en que las cadenas interaccionan entre sí además de tener cierto grado de estruc-

tura terciaria. Los contactos intermoleculares se forman, principalmente, entre las hebras. Sin dejar de interaccionar entre sí, las cadenas adquieren, cada una por separado, la estructura terciaria nativa. De esta manera, aparecen conformaciones en que una proteína está plegada y la otra no, pero interaccionando la una con la otra. El número de contactos inter-cadena de esta zona del mapa es moderadamente alto (entre 10 y 30) y el número de contactos intramoleculares por cadena, intermedio (entre 45 y 75 aproximadamente). A esta zona del mapa también corresponden conformaciones en que las cadenas tienen un grado elevado de estructura terciaria, pero no la nativa, y están interaccionando entre sí.

- Por último, a valores elevados de contactos intra-cadena aparecen configuraciones en que las dos cadenas están en el estado plegado. El número de contactos intermoleculares es distinto de cero, pero ya hemos dicho que esto es un efecto del potencial, y no lo consideramos propiamente un proceso de agregación.

En los histogramas de contactos de los sistemas de una sola cadena de la Figura 5.21 a T_m^* podemos ver que la probabilidad de las configuraciones con estructura intermedia entre el estado plegado y el desnaturalizado, para 2GB1, es muy distinta de cero, debido a que la barrera de energía es muy pequeña. Cuando la concentración es alta, las interacciones intermoleculares provocan que estas configuraciones se estabilicen y que la transición de unos estados a otros se convierta en un proceso gradual a través de ellas. Este efecto de la concentración ya lo hemos observado también para 1W4E con el potencial de interacciones locales modificadas.

En el caso de 1YPB, la desaparición de la barrera se produce de la misma manera. Además, si nos fijamos en el histograma de contactos terciarios formados a la temperatura de transición del sistema de una cadena de la Figura 5.21, vemos que el estado plegado presenta una variabilidad conformacional elevada. El pico que corresponde al estado plegado es más ancho que el que corresponde al estado desnaturalizado. Con este potencial, la transición de 1YPB consiste en el colapso de la cadena. Al dismi-

nir la temperatura, se adquiere la estructura terciaria de forma gradual. El colapso es una transición de primer orden, es decir, presenta una barrera de energía libre, pero la población de estados de energía intermedia a T_m^* es significativa, de manera que al aumentar la concentración, estos estados sufren un grado de estabilización tal que la barrera desaparece. Estas observaciones justifican, en buena medida, que las transiciones de plegamiento que presentan barreras de energía tan bajas puedan considerarse procesos graduales de tipo *downhill*.

Al contrario que para 2GB1 y 1YPB, la transición de 1W4E sigue siendo de primer orden cuando se pasa a un sistema de dos cadenas. Ya hemos comentado en el capítulo anterior y a lo largo de éste que la topología de 1W4E favorece enormemente la cooperatividad del tránsito de plegamiento. En la Figura 5.21 podemos distinguir varios estados bien diferenciados a lo largo de la transición. A valores bajos de contactos intramoleculares formados aparece el estado desnaturalizado. En este estado se aprecia que existe bastante interacción intermolecular. Por otra parte, el estado con un número de contactos intra-cadena entre 50 y 60, corresponde al plegamiento de las dos cadenas del sistema. A valores intermedios entre el estado desnaturalizado y el de las dos cadenas en el estado plegado aparece un tercer estado formado por conformaciones en que una cadena está plegada y la otra en el estado desnaturalizado. Sin embargo, a diferencia de lo que ocurre con el potencial sin modificar, este estado incluye también conformaciones en que ambas cadenas presentan estructura terciaria parcial y están interaccionando entre sí. Es decir, para 1W4E ocurre, en cierto grado, lo mismo que para 2GB1 y 1YPB: se produce la estabilización de configuraciones con energía y estructura intermedia por acción de las interacciones intermoleculares. Por último, aparece un estado en el que las cadenas tienen formados pocos contactos intramoleculares, pero muchos contactos intermoleculares. Ya hemos comentado que, debido a su estructura, el intercambio de dominios es un mecanismo poco favorecido en esta proteína, por lo que las conformaciones que dan lugar a este estado agregado corresponden a las cadenas con poca estructura

terciaria formada, pero con cierta estructura local, que interaccionan de manera poco específica entre sí.

Por último vemos que, a la temperatura de transición, para 1BBL aparece un único estado. Las características de la transición de esta proteína no difieren mucho de las que presenta con el potencial sin modificar, excepto que, en este caso, no se estabiliza ningún intermedio de plegamiento al aumentar la concentración. En las secciones anteriores ya hemos visto que, debido a la ausencia de barrera, la asociación y el plegamiento son procesos simultáneos en esta proteína.

Las transiciones de los sistemas de varias cadenas con este potencial se parecen mucho a las que tienen lugar con el potencial sin modificar. Sin embargo, la presencia del potencial de colapso hidrófobo de Zhou hace que el colapso de las cadenas sea más inespecífico. Debido a esto disminuyen las barreras entre unos estados y otros, llegando a desaparecer cuando aumenta la concentración. Tener en consideración la secuencia, incluso de manera tan genérica como lo hace el potencial de Zhou utilizado, provoca que la transición esté peor definida, dando lugar a una fuerte competencia entre agregación y plegamiento. Para saber cuál de los dos procesos está más favorecido hemos analizado las propiedades estructurales de los sistemas de dos cadenas a baja temperatura ($T_{str}^* = 0.75T_m^*$, como hasta ahora). Los resultados, que aparecen en la Tabla 5.7, muestran que, para las tres primeras proteínas, el radio de giro de las cadenas es igual al del estado nativo. Por otra parte, el valor medio de *rmsd* respecto a la estructura nativa es también, para estas proteínas, muy bajo, indicando que el proceso más favorecido es el plegamiento y que las cadenas adquieren esencialmente la estructura del estado nativo.

Sin embargo, la estructura de cada cadena de 1YPB a la temperatura T_{str}^* presenta desviaciones importantes respecto a la estructura nativa, como indica el mayor valor de *rmsd*. Esto se debe a que, a esta temperatura, aparecen tanto configuraciones en las que las dos cadenas están plegadas como configuraciones en las que están formando un

Tabla 5.7: Valor medio del radio de giro (R_g) y de $rmsd$ respecto a la estructura nativa (con su desviación típica) de las configuraciones recogidas a una temperatura $T_{str}^* = 0.75T_m^*$ en sistemas de dos cadenas para el potencial con las interacciones de largo alcance modificadas. R_g^{PDB} es el radio de giro de la estructura nativa.

| | $R_g^{PDB}/\text{\AA}$ | $R_g/\text{\AA}$ | $rmsd/\text{\AA}$ |
|------|------------------------|------------------|-------------------|
| 1BBL | 9.3 | 9.0 ± 0.3 | 2 ± 1 |
| 1W4E | 9.0 | 9.0 ± 0.2 | 2.0 ± 0.8 |
| 2GB1 | 10.1 | 10.12 ± 0.08 | 0.8 ± 0.2 |
| 1YPB | 10.8 | 11.2 ± 0.6 | 3 ± 4 |

agregado por intercambio de dominios. Este intercambio de dominios, que podemos ver esquematizado en la Figura 5.23, se produce por el intercambio de la última horquilla β , lo cual no supone un aumento demasiado elevado en el radio de giro, pero sí en el valor de $rmsd$.

Si bien en el caso de 1YPB la agregación está muy favorecida, de manera que, a baja temperatura, aparecen tanto estados agregados como el estado plegado, en el resto de los casos la frustración sigue siendo insuficiente y prevalece el proceso de plegamiento. Sin embargo, tomar en consideración la secuencia, aunque sea sólo de forma aproximada, produce cambios considerables en la transición que dan lugar a que la competencia entre agregación y plegamiento sea importante. De hecho, el plegamiento se produce a través de estados agregados, sobre todo en el caso de 1BBL, 2GB1 y 1YPB, debido a la baja (o nula) barrera de energía libre. Las interacciones entran en conflicto unas con otras de tal manera que se estabilizan estados con energía y estructura intermedia entre el estado plegado y el desnaturalizado, precisamente por el efecto de las interacciones intermoleculares. Por otra parte, aunque es cierto que la presencia de una barrera de energía libre es un mecanismo importante para impedir la agregación, la secuencia juega también un papel importante, lo cual puede explicar el comportamiento diferente de

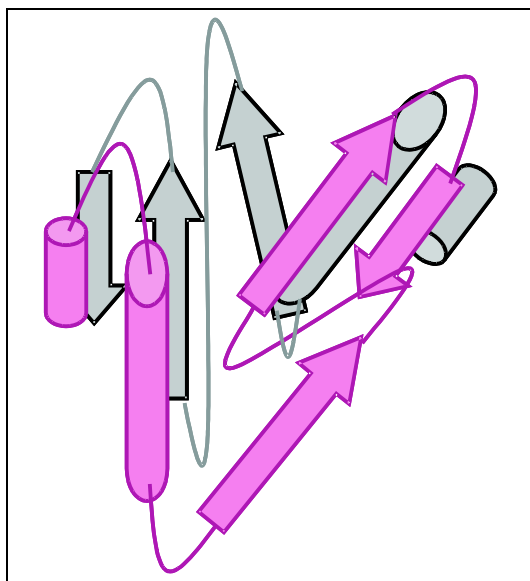


Figura 5.23: Representación esquemática del agregado que aparece a temperatura $T_{str}^* = 0.75T_m^*$ por intercambio de dominios en 1YPB.

2GB1 y 1YPB.

5.4. Resumen y conclusiones del Capítulo.

En este último capítulo hemos pasado a aplicar un potencial basado en la topología del estado nativo al proceso de agregación a través de la simulación de sistemas de varias cadenas. De esta manera, hemos podido comprobar cuál es el límite de aplicabilidad, en principio reducido, de este tipo de potenciales. Además, hemos modificado el potencial de manera que las fuentes de frustración en el sistema sean distintas. Por un lado, hemos aumentado la flexibilidad del sistema definiendo un potencial de interacción genérico para las interacciones locales. En segundo lugar, hemos introducido parcialmente el efecto de la secuencia a modo de perturbación dentro del esquema de interacciones de largo alcance. Aunque las modificaciones introducidas en el potencial no describen de manera adecuada la transición en algunos casos, el conjunto de resultados obtenidos nos ha permitido establecer relaciones entre las propiedades de la transición y la mayor o menor tendencia a la formación de agregados.

En primer lugar, hemos podido comprobar que la presencia o ausencia de barrera de energía libre en la transición de plegamiento tiene un efecto directo en el la competición entre los procesos de agregación y plegamiento que estamos estudiando. Con el potencial basado únicamente en la topología del estado nativo hemos podido comprobar cómo la ausencia de esta barrera provoca que a lo largo de la transición de plegamiento intervengan las interacciones intermoleculares de manera representativa. Además, con este potencial, en el caso de proteínas cuyo plegamiento es de dos estados, la formación del estado nativo está claramente favorecida.

Por otra parte, la altura de esta barrera de energía libre también tiene una influencia directa en este equilibrio. Si la barrera es tal que la población de estados de energía intermedia entre el plegado y el desnaturalizado es muy distinta de cero, al aumentar la concentración de proteína, las conformaciones con esta energía intermedia pueden interaccionar entre sí. Si el grado de frustración es lo suficientemente elevado, esta interacción provoca la estabilización de estos estados intermedios, de manera que la agregación cobra importancia a lo largo de la transición de plegamiento. Este efecto lo hemos observado cuando hemos introducido el efecto de la secuencia y hemos visto que, dependiendo de ella, se favorece un proceso u otro (para 1YPB con este potencial aparecen agregados a baja temperatura, mientras que las otras proteínas dan lugar al estado nativo). Hemos observado efectos similares en el caso de 1W4E con el potencial genérico de interacciones locales o para sistemas de cuatro cadenas de 2GB1 con el potencial cruzado basado en la topología del estado nativo. Que la población de estados con energía intermedia entre el estado plegado y el desnaturalizado sea lo suficientemente significativa como para verse estabilizados por la acción de las interacciones inter-cadena justifica que estos procesos de plegamiento puedan considerarse procesos graduales sin barrera.

La relación entre la barrera de energía libre y la mayor o menor tendencia a la agregación que hemos observado está de acuerdo con la hipótesis de que una transición de

primer orden podría haber sido seleccionada a lo largo de la evolución como mecanismo preventivo de dicho proceso de agregación² y que las modificaciones que reducen la cooperatividad en el plegamiento favorecen la la formación de agregados⁵¹.

Si bien el grado de frustración debido al efecto de la secuencia es importante, la flexibilidad de cadena tiene un efecto mucho más drástico. De hecho, un potencial local de tipo genérico introduce un grado de frustración tal que, en todos los casos que hemos estudiado, a bajas temperaturas encontramos estados agregados. Por un lado, encontramos estados colapsados que interaccionan entre sí. Aunque no consideramos estas conformaciones como agregados propiamente dichos, tampoco se puede considerar que se trate del estado nativo ya que, debido a su flexibilidad, se produce una deformación de la estructura nativa para incrementar el número de contactos inter-cadena. Además, aparecen también agregados que se forman tanto por intercambio de dominios como por interacción inespecífica entre cadenas abiertas.

Por otra parte, la modificación de las interacciones locales nos ha permitido estudiar el comportamiento de sistemas en los que aparece un intermedio de plegamiento. En la transición de plegamiento de 2GB1 y 1YPB descrita con este potencial aparece un intermedio de plegamiento que, en el sistema de dos cadenas, está en equilibrio con estados agregados. Esto es comparable con las observaciones experimentales obtenidas a partir de mutantes de la proteína GB1. Podemos decir entonces que la presencia de intermedios de plegamiento es un factor que favorece la formación de agregados, lo cual coincide con las observaciones experimentales¹⁵⁷.

Que el proceso de agregación esté más favorecido para 2GB1 es producto de su topología. Debido a su simetría estructural, el proceso de intercambio de dominios está facilitado. De hecho, en los sistemas de cuatro cadenas estudiados con el potencial basado sólo en la topología del estado nativo, es la única de las cuatro proteínas que forma tetrámeros a lo largo de la transición. A partir de un mutante de esta proteína se ha detectado experimentalmente un tetrámero cuya estructura no se parece a la del que

aparece a lo largo de nuestras simulaciones¹⁵⁵. Sin embargo, el hecho de que tanto en la simulación como experimentalmente se forme un tetrámero puede considerarse una evidencia de la relación que existe entre el proceso de agregación y la topología del estado nativo. Además, con el potencial original se pueden reproducir ciertas características experimentales como que, a lo largo del plegamiento, aparezcan agregados que se forman a partir del estado desnaturalizado y que están en equilibrio con el estado plegado. Podemos decir entonces que ciertos aspectos de la competición entre los procesos de agregación y plegamiento están relacionados con el principio de frustración mínima.

Como ya hemos dicho, para poder estudiar procesos de agregación más generales es necesario recurrir a potenciales más genéricos. Por un lado, tomar en consideración la secuencia puede darnos información sobre la mayor o menor tendencia de una proteína concreta a formar agregados. Por otra parte, hemos visto que describir un potencial genérico de interacción local hace que la formación de agregados esté ligeramente favorecida en todos los casos. Esto tiene una mayor relación con la observación experimental de que todas las proteínas tienen cierta tendencia a formar agregados independientemente de su estructura y secuencia si la concentración es lo suficientemente elevada^{2,46}. En resumen podemos decir que, aunque los potenciales basados en la topología del estado nativo son una herramienta potente para estudiar el plegamiento de proteínas, es necesaria una descripción de las interacciones de carácter más general para poder estudiar con mayor versatilidad el proceso de agregación.

Capítulo 6

Resumen y conclusiones generales de esta Tesis.

A lo largo de esta Tesis hemos estudiado cómo definir correctamente potenciales de interacción basados en la topología del estado nativo de las proteínas y cuáles son sus aplicaciones. Hemos comprobado que resultan de gran utilidad para estudiar las características del proceso de plegamiento y que pueden ser útiles para un primer acercamiento al estudio de procesos de agregación, aunque para este último objetivo resultan en general insuficientes.

En el primer Capítulo hemos planteado los tres objetivos principales de esta Tesis, que son el estudio metodológico de este tipo de potenciales, comprobando cómo los distintos parámetros influyen en las propiedades de la transición que se define y buscando los más adecuados. Además, es importante destacar que estos potenciales dependen, directamente, de la calidad de los datos experimentales respecto a la estructura de proteínas. Es decir, para poder estudiar de manera apropiada el proceso de plegamiento es necesario partir de la estructura nativa considerada como “real”. Por tanto, tomamos las estructuras determinadas por rayos X o RMN como válidas, no sin comprobar cuáles son las implicaciones de utilizar el mapa que resulta de la estructura determinada por cada

una de estas técnicas. Por otra parte, hemos utilizado este tipo de potenciales para estudiar cómo intervienen los diferentes tipos de interacciones en el proceso de plegamiento y cuáles son las características topológicas que determinan las propiedades termodinámicas del mismo. Finalmente, hemos ampliado y modificado el potencial para estudiar sistemas de varias cadenas y poder así comprobar su aplicabilidad al estudio de procesos de agregación. Para llevar a cabo estos estudios hemos utilizado las herramientas de la química computacional. Hemos realizado simulaciones de los sistemas que nos interesan utilizando el método de Monte Carlo, combinado con la técnica de *parallel tempering*, explicado en el Capítulo 2. En dicho capítulo también hay una descripción del método WHAM y demás tipos de tratamientos de datos que hemos realizado para analizar los resultados de las simulaciones.

A lo largo de los tres Capítulos centrales del presente trabajo (3, 4 y 5), hemos dado respuestas a las preguntas que plantean los objetivos propuestos. En primer lugar, hemos visto que la definición de los parámetros del potencial basado en topología que estamos utilizando es sumamente importante para poder tener una descripción adecuada del proceso de plegamiento. Por un lado, una definición adecuada del mapa de contactos garantiza el correcto funcionamiento del potencial, dado que el primero es la base del segundo. Hemos visto que la mejor manera de definir este mapa de contactos en un modelo reducido es a partir del empaquetamiento de todos los átomos pesados de la proteína. Es decir, podemos recurrir a una descripción detallada de la proteína para definir el mapa de contactos y aplicarlo a un modelo sencillo para realizar las simulaciones. Hemos visto que este modelo sencillo es suficiente para reproducir los datos experimentales, al menos cualitativamente. Por otra parte, independientemente de la función de potencial que se esté utilizando, la anchura del potencial es un parámetro de suma importancia, ya que determina la presencia o no de una barrera de energía libre y, además, determina las propiedades termodinámicas, no sólo del estado nativo, sino también, e incluso en mayor medida, del estado desnaturalizado. En la última parte del Capítulo 3, hemos

visto también que, una vez definido correctamente el mapa de contactos, éste tiene una gran influencia en las propiedades del tránsito de plegamiento. Por comparación de la estructura de una misma proteína determinada mediante dos técnicas experimentales diferentes hemos visto que, aunque las características generales del proceso de plegamiento vienen determinadas por la topología general del estado nativo, la presencia de contactos concretos puede producir cambios en la altura de la barrera o la definición del estado nativo que existe a la temperatura de transición.

En segundo lugar, en el Capítulo 4 hemos comprobado cuál es el papel de los dos tipos de interacciones que estamos considerando, que son las interacciones locales y las de largo alcance. La principal conclusión de este capítulo es que, para que el plegamiento presente las características de una transición cooperativa, también llamada de tipo “todo o nada”, es necesario que exista un equilibrio razonable entre ambos tipos de interacciones. Este equilibrio en las interacciones se traduce en un balance entre energía y entropía que da lugar a la aparición de una barrera de energía libre entre el estado nativo y el desnaturalizado. Además, hemos visto que este equilibrio entre las interacciones no se refiere tanto a la intensidad relativa de las mismas, sino al mapa de contactos que define el estado nativo. Es decir, la distribución de contactos locales y de largo alcance en el mapa de contactos del estado nativo es lo que determina la presencia o no de barrera de energía libre. Un mejor empaquetamiento de la estructura terciaria y una distribución homogénea de los contactos a lo largo de los distintos valores de orden de contacto son factores estructurales que favorecen un plegamiento de dos estados. A partir del estudio de las características estructurales de un conjunto amplio de proteínas hemos podido comprobar que estas características son generales a la mayoría de las proteínas, mientras que un comportamiento de tipo *downhill*, es decir, un plegamiento sin barrera, se puede explicar a partir de propiedades estructurales diferentes. De hecho, hemos observado que hay proteínas pequeñas con una proporción de contactos de largo alcance pequeña que presentan con nuestro modelo un plegamiento sin barrera, en concordancia con los datos

experimentales de la bibliografía. Finalmente, hemos visto que no es fácil establecer un límite claro entre el tipo de estructura que da lugar a un tipo de plegamiento y otro, ya que la presencia de contactos concretos es frecuentemente la que determina las características de la transición termodinámica.

Por último, en el Capítulo 5, hemos utilizado el potencial basado en la topología del estado nativo para estudiar su comportamiento en sistemas de varias cadenas. Este estudio nos ha permitido comprobar la relación entre el principio de frustración mínima y el proceso de agregación, dado que el potencial es capaz de reproducir procesos como la formación de estados agregados, en equilibrio con el estado desnaturalizado, que se forman en el proceso de plegamiento. Por otra parte, hemos visto cómo la presencia de una barrera de energía libre en el plegamiento ciertamente dificulta la formación de agregados, siempre y cuando sea lo suficientemente grande. Además, hemos comprobado que existe una relación entre la simetría estructural del estado nativo y la tendencia a formar agregados por intercambio de dominios. Sin embargo, dadas las características del potencial que, inevitablemente, conduce al estado nativo mediante un muestreo adecuado como el nuestro, no es totalmente posible estudiar las características del proceso de agregación. Para poder hacerlo, es necesario definir interacciones de carácter más genérico. De hecho, modificando el potencial basado en la topología hemos visto que introducir modificaciones en el esquema de interacciones de largo alcance, de forma que se considere la secuencia de manera aproximada, aumenta la tendencia a la agregación. En el caso de modificar las interacciones locales para definir un potencial de corto alcance más genérico, esta tendencia es aún mayor.

En resumen, en este trabajo se pone de manifiesto la validez del principio de frustración mínima para el plegamiento de proteínas. Podemos explicar, por tanto, características del proceso de plegamiento observadas experimentalmente a partir de la correcta consideración de la topología del estado nativo. En esta Tesis nos hemos centrado en las propiedades termodinámicas del proceso, pero también es posible realizar estudios sobre

algunas características dinámicas del proceso y los caminos de plegamiento. Además, con la metodología y modelo propuestos hemos podido estudiar la aparición de intermedios de plegamiento. Todo ello pone de manifiesto la trascendencia del modelo desarrollado en este trabajo, y su posible aplicación a otros problemas de mayor relevancia biológica.

Bibliografía

1. C. Gómez-Moreno y J. Sancho, *Estructura de Proteínas*. Ariel, Barcelona, 2003.
2. C.M. Dobson, “The structural basis of protein folding and its links with human disease”, *Phil. Trans. R. Soc. Lond. B* **356**: 133, 2001.
3. T.E. Creighton, *Proteins. Structures and Molecular properties*. 2ª ed., Freeman & Co., Nueva York, 1993.
4. C.B. Anfinsen y H.A. Scheraga, “Experimental and theoretical aspects of protein folding”, *Adv. Protein Chem.* **29**: 205, 1975.
5. K.A. Dill, S.B. Ozkan, T.R. Weikl, J.D. Chodera y V.A. Voelz, “The protein folding problem: when will it be solved?”, *Curr. Opin. Struct. Biol.* **17**: 342, 2007.
6. C. Levinthal, “Are there pathways for protein folding?”, *J. Chim. Phys.* **65**: 44, 1968.
7. J.D. Bryngelson y P.G. Wolynes, “Spin glasses and the statistical mechanics of protein folding”, *Proc. Natl. Acad. Sci. USA* **84**: 7524, 1987.
8. J.N. Onuchic, Z. Luthey-Schulten y P.G. Wolynes, “Theory of protein folding: The energy landscape perspective”, *Annu. Rev. Phys. Chem.* **48**: 545, 1997.
9. J.N. Onuchic y P.G. Wolynes, “Theory of protein folding”, *Curr. Opin. Struct. Biol.* **14**: 70, 2004.
10. S.S. Plotkin, J. Wang y P.G. Wolynes, “Statistical mechanics of a correlated energy landscape model for protein folding funnels”, *J. Chem. Phys.* **106**: 2932, 1997.
11. S.E. Jackson, “How do small single-domain proteins fold?”, *Fold. Des.* **3**: R81, 1998.
12. B. Ibarra-Molero y J.M. Sanchez-Ruiz, “Differential scanning calorimetry of proteins: An overview and some recent developments”, en *Advanced Techniques in Biophysics*, Eds. J.L.R. Arrondo y A. Alonso, Springer, Berlín, 2006.

13. B. Ibarra-Molero y J.M. Sanchez-Ruiz, "Differential scanning calorimetry: The statistical thermodynamics analysis of protein folding/unfolding", en *Protein Folding and Misfolding*, Ed. V. Muñoz, Royal Society of Chemistry, Londres, 2007.
14. M.M. García-Mira, M. Sadqi, N. Fisher, J.M. Sanchez-Ruiz y V. Muñoz, "Experimental identification of downhill protein folding", *Science* **298**: 2191, 2002.
15. F.Y. Oliva y V. Muñoz, "A simple thermodynamic test to discriminate between two-state and downhill folding", *J. Am. Chem. Soc.* **126**: 8596, 2004.
16. V. Muñoz y J.M. Sanchez-Ruiz, "Exploring Protein-Folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments", *Proc. Natl. Acad. Sci. USA* **101**: 17646, 2004.
17. A.N. Naganathan, R. Pérez-Jiménez, J.M. Sanchez-Ruiz y V. Muñoz, "Robustness of downhill folding: Guidelines for the analysis of equilibrium folding experiments on small proteins", *Biochemistry* **44**: 7435, 2005.
18. J.M. Sturtevant, "Biochemical applications of differential scanning calorimetry", *Annu. Rev. Phys. Chem.* **38**: 463, 1987.
19. Y. Zhou, C.K. Hall y M. Karplus, "The calorimetric criterion for a two-state process revisited", *Protein Sci.* **8**: 1064, 1999.
20. M. Gruebele, "Downhill protein folding: Evolution meets physics", *C.R. Biol* **328**: 701, 2005.
21. W. Eaton, "Searching for 'downhill scenarios' in protein folding", *Proc. Natl. Acad. Sci. USA* **96**: 5897, 1999.
22. A. Akmal y V. Muñoz, "The nature of the free energy barriers to two-state folding", *Proteins* **57**: 142, 2004.
23. H.K. Nakamura, M. Sasai y M. Takano, "Squeezed exponential kinetics to describe a nonglassy downhill folding as observed in a lattice protein model", *Proteins* **55**: 99, 2004.
24. A.N. Naganathan y V. Muñoz, "Scaling of folding times with protein size", *J. Am. Chem. Soc.* **127**: 480, 2005.
25. A.N. Naganathan, J.M. Sanchez-Ruiz y V. Muñoz, "Direct measurement of barrier heights in protein folding", *J. Am. Chem. Soc.* **127**: 17970, 2005.

26. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov y P.E. Bourne, "The Protein Data Bank", *Nucleic Acids Res.* **28**: 235, 2000.
27. M. Sadqi, D. Fushman y V. Muñoz, "Atom-by-atom analysis of global downhill protein folding", *Nature* **442**: 04859, 2006.
28. N. Ferguson, P.J. Schartau, T.D. Sharpe, S. Sao y A.R. Fersht, "One-state downhill *versus* conventional protein folding", *J. Mol. Biol.* **344**: 295, 2004.
29. N. Ferguson, R. Day, C.M. Johnson, M.D. Allen, V. Dagget y A.R. Fersht, "Simulation and experiment at high temperatures: Ultrafast folding of a thermophilic protein by nucleation-condensation", *J. Mol. Biol.* **347**: 855, 2005.
30. N. Ferguson, T.D. Sharpe, P.J. Schartau, S. Sato, M.D. Allen, C.M. Johnson, T.J. Rutherford y A.R. Fersht, "Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family", *J. Mol. Biol.* **353**: 427, 2005.
31. N. Ferguson, T.D. Sharpe, C.M. Johnson y A.R. Fersht, "The transition state for folding of a peripheral subunit-binding domain contains robust and ionic-strength dependent characteristics", *J. Mol. Biol.* **356**: 1237, 2006.
32. N. Ferguson, C.M. Johnson, P.J. Schartau y A.R. Fersht, "Analysis of 'downhill' protein folding", *Nature* **445**: E14, 2007.
33. Z. Zhou e Y. Bai, "Analysis of protein-folding cooperativity", *Nature* **445**: E16, 2007.
34. M. Sadqi, D. Fishman y V. Muñoz, "Sadqi *et. al.* reply", *Nature* **445**: E17, 2007.
35. D.J. Brockwell, D.A. Smith y S.E. Radford, "Protein folding mechanisms: New methods and emerging ideas", *Curr. Opin. Struct. Biol.* **10**: 16, 2000.
36. S.S. Plotkin, "Speeding protein folding beyond the Gō model: How a little frustration sometimes helps", *Proteins* **45**: 337, 2001.
37. A.R. Fersht, "Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism", *Proc. Natl. Acad. Sci. USA* **97**: 1525, 2000.
38. H.S. Chan, "Matching speed and locality", *Nature* **392**: 761, 1998.
39. B. Nölting, "Protein Folding Kinetics", Springer, Nueva York, 2006.

40. K.W. Plaxco, K.T. Simons y D. Baker, "Contact order, transition state placement and the refolding rates of single domain proteins", *J. Mol. Biol.* **277**: 985, 1998.
41. L.A. Munishkina, E.M. Cooper, V.N. Uversky y A.L. Fink, "The effect of macromolecular crowding on protein aggregation and amyloid fibril formation", *J. Mol. Rec.* **17**: 456, 2004.
42. M. Sunde, L.C. Serpell, M. Bartlam, P.E. Fraser, M.B. Pepys y C.C.F. Blake, "Common core structure of amyloid fibrils by synchrotron X-ray diffraction", *J. Mol. Biol.* **273**: 729, 1997.
43. J.I. Guijarro, M. Sunde, J.A. Jones, I.D. Campbell y C.M. Dobson, "Amyloid fibril formation by an SH3 domain", *Proc. Natl. Acad. Sci. USA* **95**: 4224, 1998.
44. J.W. Kelly, "Alternative conformations of amyloidogenic proteins govern their behavior", *Curr. Opin. Struct. Biol.* **6**: 11, 1996.
45. F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi y C.M. Dobson, "Designing conditions for *in vitro* formation of amyloid protofilaments and fibrils", *Proc. Natl. Acad. Sci. USA* **96**: 3590, 1999.
46. E.Y. Chi, S. Krishnan, T. W. Randolph y J.F. Carpenter, "Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation", *Pharm. Res.* **20**: 1325, 2003.
47. V.N. Uversky, D.J. Segel, S. Doniach y A.L. Fink, "Association-induced folding of globular proteins", *Proc. Natl. Acad. Sci. USA* **95**: 5480, 1998.
48. M. Silow, Y.J. Tan, A.R. Fersht y M. Oliveberg, "Formation of short-lived protein aggregates directly from the coil in two-state folding", *Biochemistry* **98**: 13006, 1999.
49. Y. Liu, G. Gotte, M. Libonati y D. Eisenberg, "A domain-swapped RNase A dimer with implications for amyloid formation", *Nat. Struct. Biol.* **8**: 211, 2001.
50. M. Vendruscolo y C.M. Dobson, "Towards complete descriptions of the free-energy landscapes of proteins", *Phil. Trans. R. Soc. A.* **363**: 433, 2005.
51. Q. Hua y M.A. Weiss, "Mechanism of insulin fibrillation", *J. Biol. Chem.* **279**: 21449, 2004.
52. R.A. Friesner Ed., *Computational methods for protein folding* Wiley, Nueva York, 2001.

53. H. Taketomi, Y. Ueda y N. Gō, "Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions", *Int. J. Pept. Protein Res.* **7**: 445, 1975.
54. N. Gō y H. Taketomi, "Respective roles of short- and long-range interactions in protein folding", *Proc. Natl. Acad. Sci. USA* **75**: 559, 1978.
55. N. Gō, "Protein folding as a stochastic process", *J. Stat. Phys.* **30**: 413, 1983.
56. J. Shimada, E.L. Kussel y E.I. Shakhnovich, "The folding thermodynamics and kinetics of cambrin using an all-atom Monte Carlo simulation", *J. Mol. Biol.* **308**: 79, 2001.
57. A. Linhananta y Y. Zhou, "The role of sidechain packing and native contact interactions in folding: Discontinuous molecular dynamics folding simulations of an all-atom Gō model of fragment B of *Staphylococcal* protein A", *J. Chem. Phys.* **117**: 8983, 2002.
58. J. Wang y W. Wang, "Folding transition of model protein chains characterized by partition function zeros", *J. Chem. Phys.* **118**: 2952, 2003.
59. C. Clementi, A.E. García y J.N. Onuchic, "Interplay among tertiary contacts, secondary structure formation mechanism: All-atom representation study of protein L", *J. Mol. Biol.* **326**: 993, 2003.
60. C. Clementi, H. Nymeyer y J.N. Onuchic, "Topological and energetic factors: What determines the structural details of the transition state ensemble and 'En-route' intermediates for protein folding? An investigation for small globular proteins", *J. Mol. Biol.* **298**: 937, 2000.
61. C. Clementi, P.A. Jennings y J.N. Onuchic, "How native state topology affects the folding of dihydrofolate reductase and interleukin-beta", *Proc. Natl. Acad. Sci. USA* **97**: 5871, 2000.
62. V.S. Pande y D.S. Roskhar, "Is the molten globule a third phase of proteins?", *Proc. Natl. Acad. Sci. USA* **95**: 1490, 1998.
63. D. Bratko y H.W. Blanch, "Competition between protein folding and aggregation: A three-dimensional lattice simulation", *J. Chem. Phys.* **114**: 561, 2001.
64. N. Combe y D. Frenkel, "Phase behavior of a lattice protein model", *J. Chem. Phys.* **118**: 9015, 2003.

65. J.E. Shea, J.N. Onuchic y C.L. Brooks III, "Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A", *Proc. Natl. Acad. Sci. USA* **96**: 12512, 1999.
66. B. Erman, "Analysis of multiple folding routes of proteins by a coarse-grained dynamics model", *Biophys. J.* **81**: 2534, 2001.
67. M. Cieplak y T.X. Hoang, "The range of the contact interactions and the kinetics of the Gō models of proteins", *Int. J. Mod. Phys. C* **13**: 1231, 2002.
68. J. Shimada y E.I. Shakhnovich, "The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation", *Proc. Natl. Acad. Sci. USA* **99**: 11175, 2002.
69. W. Humphrey, A. Dalke y K. Schulten, "VMD-Visual Molecular Dynamics", *J. Mol. Graph.* **14**: 33, 1996.
70. M. Cieplak y T.X. Hoang, "Folding of proteins in Gō models with angular interactions", *Physica A* **330**: 195, 2003.
71. Y. Zhou y M. Karplus, "Folding thermodynamics of a model three-helix-bundle protein", *Proc. Natl. Acad. Sci. USA* **94**: 14429, 1997.
72. Y. Zhou y M. Karplus, "Folding of a model three-helix bundle protein: A thermodynamic and kinetic analysis", *J. Mol. Biol.* **293**: 917, 1999.
73. Y. Zhou y A. Linhananta, "Thermodynamics of an all-atom off-lattice model of the fragment B of Staphylococcal protein A: Implication for the origin of the cooperativity of protein folding", *J. Phys. Chem. B* **106**: 1481, 2002.
74. H. Jang, C.K. Hall y Y. Zhou, "Folding thermodynamics of model four-strand antiparallel β -sheet proteins", *Biophys. J.* **82**: 646, 2002.
75. H. Kaya y H.S. Chan, "Solvation effects and driving forces for protein thermodynamics and kinetic cooperativity: How adequate is native-centric topological modeling?", *J. Mol. Biol.* **326**: 911, 2003.
76. M. Knott, H. Kaya y H.S. Chan, "Energetics of protein thermodynamic cooperativity: contributions of local and nonlocal interactions", *Polymer* **45**: 623, 2004.
77. K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas y H.S. Chan, "Principles of protein folding - A perspective from simple exact models", *Protein Sci.* **4**: 561, 1995.

78. R. Unger y J. Moult, "Local interactions dominate folding in a simple protein model", *J. Mol. Biol.* **259**: 988, 1996.
79. A. Irback, C. Peterson, F. Potthast y O. Sommelius, "Local interactions and protein folding: A three-dimensional off-lattice approach", *J. Chem. Phys.* **107**: 273, 1997.
80. S. Govindarajan y R.A. Goldstein, "Optimal local propensities for model proteins", *Proteins* **22**: 413, 1995.
81. V.I. Abkevich, A.M. Gutin y E.I. Shakhnovich, "Impact of local and non-local interactions on thermodynamics and kinetics of protein folding", *J. Mol. Biol.* **252**: 460, 1995.
82. B. Gillespie y K.W. Plaxco, "Nonglassy kinetics in the folding of a simple single-domain protein", *Proc. Natl. Acad. Sci. USA* **97**: 12014, 2000.
83. J.E. Shea, J.N. Onuchic y C.L. Brooks III, "Probing the folding free energy landscape of the src-SH3 protein domain", *Proc. Natl. Acad. Sci. USA* **99**: 16064, 2002.
84. A.I. Jewett, V.S. Pande y K.N. Plaxco, "Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates", *J. Mol. Biol.* **326**: 247, 2003.
85. P.F.N. Faisca y M.M. Telo da Gama, "Folding of small proteins: a matter of geometry?", *Mol. Phys.* **103**: 2903, 2005.
86. R. Zwanzig, "Simple model of protein folding kinetics", *Proc. Natl. Acad. Sci. USA* **92**: 9801, 1995.
87. R. Doyle, R. Simons, H. Qian y D. Baker, "Local interactions and the optimization of protein folding", *Proteins* **29**: 282, 1997.
88. M.M. Gromiha y S. Selvaraj, "Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: Application of long-range order to folding rate prediction", *J. Mol. Biol.* **310**: 27, 2001.
89. H. Kaya y H.S. Chan, "Contact order dependent protein folding rates: Kinetic consequences of a cooperative interplay between favorable nonlocal interactions and local conformational preferences", *Proteins* **52**: 524, 2003.
90. H.S. Chan, S. Shimizu y H. Kaya, "Cooperativity principles in protein folding", *Methods Enzymol.* **380**: 350, 2004.

91. P.F.N. Faisca, M.M. Telo da Gamma y R.C. Ball, "Folding and form: Insights from lattice simulations", *Phys. Rev. E* **69**: 1917, 2004.
92. P.F.N. Faisca y M.M. Telo da Gama, "The Gō model revisited. Native structure and the geometric coupling between local and long-range contacts", *Proteins* **60**: 712, 2005.
93. P.F.N. Faisca y M.M. Telo da Gama, "Native geometry and the dynamics of protein folding", *Biophys. Chem.* **115**: 169, 2005.
94. M. Punta y B. Rost, "Protein folding rates estimated from contact predictions", *J. Mol. Biol.* **348**: 507, 2005.
95. G. Zuo, J. Wang y W. Wang, "Folding with downhill behavior and low cooperativity of proteins", *Proteins* **63**: 165, 2006.
96. F. Ding, N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley and E.I. Shakhnovich, "Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism", *J. Mol. Biol.* **324**: 851, 2002.
97. N.L. Fawzi, V. Chubukov, L.A. Clark, S. Brown y T. Head-Gordon, "Influence of denatured and intermediate states of folding on protein aggregation", *Protein Sci.* **14**: 993, 2005.
98. A.V. Smith y C.K. Hall, "Protein refolding *versus* aggregation: Computer simulations on an intermediate-resolution protein model", *J. Mol. Biol.* **312**: 187, 2001.
99. B. Ma y R. Nussinov, "Stabilities and conformations of Alzheimer's beta-amyloid peptide oligomers (Abeta 16-22, Abeta 16-35, and Abeta 10-35): Sequence effects", *Proc. Natl. Acad. Sci. USA* **99**: 14126, 2002.
100. E. Paci, J. Gsponer, X. Salvatella y M. Vendruscolo, "Molecular dynamics studies of the process of amyloid aggregation of peptide fragments of transthyretin", *J. Mol. Biol.* **340**: 555, 2004.
101. T. Cellmer, D. Bratko, J.M. Prausnitz y H. Blanch, "The competition between protein folding and aggregation: Off-lattice minimalist model studies", *Biotech. Bioeng.* **89**: 78, 2004.
102. M. Cecchini, F. Rao, M. Seeber y A. Caflish, "Replica exchange molecular dynamics of amyloid peptide aggregation", *J. Chem. Phys.* **121**: 10748, 2004.

103. H.D. Nguyen y C.K. Hall, "Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides", *Proc. Natl. Acad. Sci. USA* **101**: 16180, 2004.
104. E. Malolepsza, M. Boniecki, A. Kolinski y L. Piela, "Theoretical model of prion propagation: A misfolded protein induces misfolding", *Proc. Natl. Acad. Sci. USA* **102**: 7835, 2005.
105. D. Bratko, T. Cellmer, J.M. Prausnitz y H.W. Blanch, "Effect of single-point sequence alteration propensity of a model protein", *J. Am. Chem. Soc.* **128**: 1683, 2006.
106. J.H. Meinke y U.H.E. Hansmann, "Aggregation of β -amyloid fragments", *J. Chem. Phys.* **126**: 014706, 2007.
107. A. Trovato, A. Maritan y F. Seno, "Aggregation of natively folded proteins: A theoretical approach", *J. Phys.: Cond. Matter* **19**: 285221, 2007.
108. L.A. Clark, "Protein aggregation determinants from a simplified model: Cooperative folders resist aggregation", *Protein Sci.* **14**: 653, 2005.
109. D. Bratko y H.W. Blanch, "Effect of secondary structure on protein aggregation: A replica exchange simulation study", *J. Chem. Phys.* **118**: 5185, 2003.
110. T. Cellmer, D. Bratko, J.M. Prausnitz y H. Blanch, "Protein-folding landscapes in multichain systems", *Proc. Natl. Acad. Sci. USA* **102**: 11692, 2005.
111. H. Jang, C.K. Hall y Y. Zhou, "Assembly and kinetic folding pathways of a tetrameric β -sheet complex: Molecular dynamics simulations on simplified off-lattice protein models", *Biophys. J.* **86**: 31, 2004.
112. S. Yang, S.S. Cho, Y. Levy, M.S. Cheung, H. Levine, P.G. Wolynes y J.N. Onuchic, "Domain swapping is a consequence of minimal frustration", *Proc. Natl. Acad. Sci. USA* **101**: 13786, 2004.
113. R.B. Dyer, "Ultrafast and downhill protein folding", *Curr. Opin. Struct. Biol.* **17**: 1, 2007.
114. A. Liwo, M. Khalili y H.A. Scheraga, "Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains", *Proc. Natl. Acad. Sci. USA* **102**: 2362, 2005.
115. U.H.E. Hansmann, "Parallel tempering algorithm for conformational studies of biological molecules", *Chem. Phys. Lett* **281**: 140, 1997.

116. Y. Sugita y Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding", *Chem. Phys. Lett.* **314**: 141, 1999.
117. M.P. Allen y D.J. Tildesley, *Computer Simulation of Liquids*. Clarendon Press, Oxford, 1989.
118. A. R. Leach, *Molecular Modelling. Principles and Applications*. Prentice-Hall, Harlow, 2001.
119. N. Rathore, M. Chopra y J. de Pablo, "Optimal allocation of replicas in parallel tempering simulations", *J. Chem. Phys.* **122**: 02411, 2005.
120. C. Ayyagari, D. Bedrov y G.D. Smith, "Equilibrium sampling of self-associating polymer solutions: A parallel selective tempering approach", *J. Chem. Phys.* **123**: 124912, 2005.
121. D. Gront, A. Kolinski y J. Skolnick, "Comparison of three Monte Carlo conformational search strategies for a proteinlike homopolymer model: Folding thermodynamics and identification of low-energy structures", *J. Chem. Phys.* **113**: 5065, 2000.
122. R. Brüschweiler, "Efficient *rmsd* measures for the comparison of two molecular ensembles", *Proteins* **50**: 26, 2003.
123. N.V. Prabhu y K.A. Sharp, "Heat capacity in proteins", *Annu. Rev. Phys. Chem.* **56**: 521, 2005.
124. A. Cooper, "Thermodynamic fluctuations in protein molecules", *Proc. Natl. Acad. Sci. USA* **73**: 2740, 1976.
125. A.M. Ferrenberg y R.H. Swendsen, "New Monte Carlo technique for studying phase transitions", *Phys. Rev. Lett.* **61**: 2635, 1988.
126. A.M. Ferrenberg y R.H. Swendsen, "Optimized Monte Carlo data analysis", *Phys. Rev. Lett.* **63**: 1195, 1989.
127. S. Kumar, D. Bouzida, R.W. Swendsen, P.A. Kollman y J.M. Rosenberg, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method", *J. Comput. Chem.* **13**: 1011, 1992.
128. J.D. Chodera, W.C. Swope, J.W. Pitera, C. Seok y K.A. Dill, "Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations", *J. Chem. Theor. and Comput.* **3**: 26, 2007.

129. L. Prieto y A. Rey, "Thermodynamic of G $\bar{5}$ -type models for protein folding", *J. Chem. Phys.* **123**: 154903, 2005.
130. A.M. Gronenborn, D.R. Filpula, N.Z. Essig, A. Achari, M. Whitlow, P.T. Wingfield y G.M. Clore, "A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G", *Science* **253**: 657, 1991.
131. P. Alexander, S. Fahnstock, T. Lee, J. Orban y P. Bryan, "Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures", *Biochemistry* **31**: 3597, 1992.
132. P. Alexander, J. Orban y P. Bryan, "Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G", *Biochemistry* **31**: 7243, 1992.
133. F.J. Blanco, M.A. Jiménez, A. Pineda, M. Rico, J. Santoro y J.L. Nieto, "NMR solution structure of the isolated N-terminal fragment of protein-G B1 domain. Evidence of trifluoroethanol induced native-like β -hairpin formation", *Biochemistry* **33**: 6004, 1994.
134. D. Idiyatullin, I. Nesmelova, V.A. Daragan y K.H. Mayo, "Heat capacities and a snapshot of the energy landscape in protein GB1 from the pre-denaturation temperature dependence of backbone NH nanosecond fluctuations", *J. Mol. Biol.* **325**: 149, 2003.
135. K. Ding, J.M. Louis y A.M. Gronenborn, "Insights into conformation and dynamics of protein GB1 during folding and unfolding by NMR", *J. Mol. Biol.* **335**: 1299, 2004.
136. E.L. McCallister, E. Alm y D. Baker, "Critical role of β -hairpin formation in protein G folding", *Nat. Struct. Biol* **7**: 669, 2000.
137. P. Derreumaux, "Role of supersecondary structural elements in protein G folding", *J. Chem. Phys.* **119**: 4940, 2003.
138. S.Y. Lee, Y. Fujitsuka, D.H. Kim y S. Takada, "Molecular simulation of protein G and alpha spectrin SH3", *Proteins* **55**: 128, 2004.
139. J. Tsai, R. Taylor, C. Chothia y M. Gerstein, "The packing density in proteins: Standard radii and volumes", *J. Mol. Biol.* **290**: 253, 1999.

140. M. Larriva, “Estudio de la influencia de la estructura en el plegamiento de proteínas”, Diploma de Estudios Avanzados, Fac. de CC. Químicas, UCM, 2007.
141. L. Prieto y A. Rey, “Influence of the native topology on the folding barrier for small proteins”, *J. Chem. Phys.* **127**: 175101, 2007.
142. L. Prieto y A. Rey, “Influence of the chain stiffness on the thermodynamics of a Gō-type model for protein folding”, *J. Chem. Phys.* **126**: 165103, 2007.
143. U. Hobohm, M. Scharf, R. Scheneider y C. Sander, “Selection of representative protein data sets”, *Protein Sci.* **1**: 409, 1992. Bases de datos actualizadas disponibles en la Web PISCES, <http://dunbrack.fccc.edu/PISCES.php>.
144. S.E. Jackson y A.R. Fersht, “Folding of chymotrypsin inhibitor 2. 1.Evidence for a two-state transition”, *Biochemistry* **30**: 10428, 1991.
145. J. Kubelka, W.A. Eaton y J. Hofrichter, “Experimental tests of villin subdomain folding simulations”, *J. Mol. Biol.* **329**: 625, 2003.
146. M. Knott y H.S. Chan, “Criteria for downhill protein folding: Calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity”, *Proteins* **65**: 373, 2006.
147. L. Qui, S.A. Pabit, A.E. Roitberg y S.J. Hagen, “Smaller and faster: The 20-residue Trp-cage protein folds in 4 μ s”, *J. Am. Chem. Soc.* **124**: 12952, 2002.
148. J. Kubelka, T.K. Chiu, D.R. Davies, W.A. Eaton y J. Hofrichter, “Sub-microsecond protein folding”, *J. Mol. Biol.* **359**: 546, 2006.
149. A.N. Naganathan, U. Doshi, A. Fung, M. Sadqi y V. Muñoz, “Dynamics, energetics, and structure in protein folding”, *Biochemistry* **45**: 5466, 2006.
150. R.J. Ellis, “Macromolecular crowding: An important but neglected aspect of intracellular environment”, *Curr. Opin. Struct. Biol* **11**: 114, 2001.
151. A.P. Minton, “The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media”, *J. Biol. Chem.* **276**: 10577, 2001.
152. M. Oliveberg, “Alternative explanations for ‘multistate’ kinetics in protein folding: Transient aggregation and changing transition-state ensembles”, *Acc. Chem. Res.* **31**: 765, 1998.

153. R.I. Dima y D. Thirumalai, "Exploring protein aggregation and self-propagation using lattice models: Phase diagram and kinetics", *Protein Sci.* **11**: 1036, 2002.
154. I.L. Byeon, J.M. Louis y A.M. Gronenborn, "A protein contorsionist: Core mutations of GB1 that induce dimerization and domain swapping", *J. Mol. Biol.* **333**: 141, 2003.
155. M.K. Frank, F. Dyda, A. Dobrodumov y A.M. Gronenborn, "Core mutations switch monomeric protein GB1 into an intertwined tetramer", *Nat. Struct. Biol.* **9**: 877, 2002.
156. M.J. Sippl, "Calculation of conformational ensembles from potentials of mean force: An approach to the knowledge-based prediction of local structures in globular proteins", *J. Mol. Biol.* **213**: 859, 1990.
157. I.L. Byeon, J.M. Louis y A.M. Gronenborn, "A captured folding intermediate involved in dimerization and domain-swapping of GB1", *J. Mol. Biol.* **340**: 615, 2004.
158. H. Zhou y Y. Zhou, "Distance-scaled, finite ideal-gas reference state structure-derived potentials of mean force for structure selection and stability prediction", *Protein Sci.* **11**: 2714, 2002.
159. M. Larriva, D. de Sancho y A. Rey, "Evaluation of a mean field potential for protein folding with different interaction centers", *Physica A* **371**: 449, 2006.